

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:41:12 ; Search time 5825.52 Seconds
(without alignments)
1239.345 Million cell updates/sec

Title: US-09-851-138C-52

Perfect score: 149

Sequence: 1 DGINFATGNLPGCSFSIFLL.....QGFSWRHRQHWTVQDCNCIS 149

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9408497

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_epool_p/US09851138/runat_28022005_120306_21465/app_query.fasta_1.1123
-DB=GenEmbl -QWMT=fastap -SURFIX=olig.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09851138 @CGN 1.1 6331 @runat_28022005_120306_21465 -NCPU=6 -ICPU=3
-NO MWAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=1120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.ev.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	56.4	447	6 A50396	A50396 Sequence 51
2	84	56.4	447	6 ARI27536	ARI27536 Sequence
3	84	56.4	447	14 HPCCOREEAL	L39317 Hepatitis C
4	29	19.5	1584	14 HPVJK070A8	D49752 Hepatitis C

5	26	17.4	447	14 HPCCOREE1R	L39297 Hepatitis C
6	26	17.4	447	14 HPCCOREE1R	L39298 Hepatitis C
7	26	17.4	447	14 HPCCOREE1R	L39312 Hepatitis C
8	26	17.4	541	6 A40613	A40613 Sequence 13
9	26	17.4	541	6 A40617	A40617 Sequence 17
10	26	17.4	541	6 A40619	A40619 Sequence 19
11	26	17.4	541	6 A40621	A40621 Sequence 21
12	26	17.4	541	6 A40623	A40623 Sequence 23
13	26	17.4	541	6 A40625	A40625 Sequence 25
14	26	17.4	541	6 A40627	A40627 Sequence 27
15	26	17.4	541	6 BD172130	BD172130 New sequ
16	26	17.4	541	6 BD172132	BD172132 New sequ
17	26	17.4	541	6 BD172133	BD172133 New sequ
18	26	17.4	541	6 BD172134	BD172134 New sequ
19	26	17.4	541	6 BD172135	BD172135 New sequ
20	26	17.4	541	6 BD172136	BD172136 New sequ
21	26	17.4	541	6 BD172137	BD172137 New sequ
22	26	17.4	541	6 AX031591	AX031591 Sequence
23	26	17.4	541	6 AX031595	AX031595 Sequence
24	26	17.4	541	6 AX031597	AX031597 Sequence
25	26	17.4	541	6 AX031599	AX031599 Sequence
26	26	17.4	541	6 AX031601	AX031601 Sequence
27	26	17.4	541	6 AX031603	AX031603 Sequence
28	26	17.4	541	6 AX031605	AX031605 Sequence
29	26	17.4	541	6 AX031861	AX031861 Sequence
30	26	17.4	541	6 AX031865	AX031865 Sequence
31	26	17.4	541	6 AX031867	AX031867 Sequence
32	26	17.4	541	6 AX031869	AX031869 Sequence
33	26	17.4	541	6 AX031871	AX031871 Sequence
34	26	17.4	541	6 AX031873	AX031873 Sequence
35	26	17.4	541	6 AX031875	AX031875 Sequence
36	26	17.4	541	6 AX032131	AX032131 Sequence
37	26	17.4	541	6 AX032135	AX032135 Sequence
38	26	17.4	541	6 AX032137	AX032137 Sequence
39	26	17.4	541	6 AX032139	AX032139 Sequence
40	26	17.4	541	6 AX032141	AX032141 Sequence
41	26	17.4	541	6 AX032143	AX032143 Sequence
42	26	17.4	541	6 AX032145	AX032145 Sequence
43	26	17.4	541	14 HPCCOREEA	D14596 Hepatitis C
44	26	17.4	541	14 HPCCOREED	D14599 Hepatitis C
45	26	17.4	541	14 HPCCOREEH	D14603 Hepatitis C

ALIGNMENTS

RESULT 1	A50396	Sequence 51 from Patent WO9613590.	447 bp	DNA	linear	PAT 07-MAR-1997
LOCUS	A50396	Sequence 51 from Patent WO9613590.				
DEFINITION	A50396					
ACCESSION	A50396					
VERSION	A50396.1	GI:2303407				
KEYWORDS						
SOURCE	unidentified					
ORGANISM	unidentified					
REFERENCE	1 (bases 1 to 447)					
AUTHORS	Maertens,G. and Stuyver,L.					
TITLE	NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS					
JOURNAL	PATENT: WO 9613590-A 51 09-MAY-1996;					
COMMENT	INNOGENETICS NV (BE)					
FEATURES	Other publication AU 3844095 960523.					
	Location/Qualifiers					
	1..447					
	/organism="unidentified"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:32644"					

ORIGIN						
Alignment Scores:						
Pred. No.:	9.26e-88	Length:	447			
Score:	84.00	Matches:	84			
Percent Similarity:	100.00%	Conservative:	0			

```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 56.38% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-52 (1-149) x A50396 (1-447)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 1 GACGGAATTAATTCGCAACAGGGAATTTACCTGGTTGCTCTTCTCTATCTTCCTCTCG 60

QY 21 AlaLeuPheSerCysLeuLeuThrProThrAlaGlyLeuGluTyrArgAsnAlaSerGly 40
Db 61 GCCTTTGTTCTCATGCTTGTACCCACAGGCTGGAGTACCGTAATGCTCCGGA 120

QY 41 LeuTyrMetValThrAsnAspCysSerAsnGlySerIleValTyrGluAlaGlyAspIle 60
Db 121 CTCTACATGTTACTAAGCACTGACGTACGTAAGGTATGATGCTGATGAGCGGGGATATT 180

QY 61 IleLeuHisLeuProGlyCysValProCysValArgSerGlyAsnThrSerArgCysTyr 80
Db 181 ATCTCCACTTACCTGGCTGTGTCCTCGGTACGCTCTGGCAATACATCAAGATGCTGG 240

QY 81 IleProValSer 84
Db 241 ATCCCTGTGAGC 252

RESULT 2
LOCUS AR127536 447 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 51 from patent US 6180768.
ACCESSION AR127536
VERSION AR127536.1 GI:14114131
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 447)
AUTHORS Maertens, G. and Stuyver, L.
TITLE Sequences of hepatitis C virus genotypes and their use as prophylactic, therapeutic and diagnostic agents
JOURNAL Patent: US 6180768-A 51 30-JAN-2001;
FEATURES
source
1. .447
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.: 9,266-88 Length: 447
Score: 84.00 Matches: 84
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 56.38% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-52 (1-149) x AR127536 (1-447)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 1 GACGGAATTAATTCGCAACAGGGAATTTACCTGGTTGCTCTTCTCTATCTTCCTCTCG 60

QY 21 AlaLeuPheSerCysLeuLeuThrProThrAlaGlyLeuGluTyrArgAsnAlaSerGly 40
Db 61 GCCTTTGTTCTCATGCTTGTACCCACAGGCTGGAGTACCGTAATGCTCCGGA 120

QY 41 LeuTyrMetValThrAsnAspCysSerAsnGlySerIleValTyrGluAlaGlyAspIle 60
Db 121 CTCTACATGTTACTAAGCACTGACGTACGTAAGGTATGATGCTGATGAGCGGGGATATT 180

QY 61 IleLeuHisLeuProGlyCysValProCysValArgSerGlyAsnThrSerArgCysTyr 80
Db 181 ATCTCCACTTACCTGGCTGTGTCCTCGGTACGCTCTGGCAATACATCAAGATGCTGG 240
```

```
QY 81 IleProValSer 84
Db 241 ATCCCTGTGAGC 252

RESULT 3
LOCUS HPCCORREAL 447 bp ss-RNA linear VRL 16-OCT-2001
DEFINITION Hepatitis C virus type 3 clone NL96 precursor protein gene, partial cds.
ACCESSION L39317
VERSION L39317.1 GI:845497
KEYWORDS Hepatitis C virus type 3
SOURCE Hepatitis C virus type 3
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
REFERENCE 1 (bases 1 to 447)
AUTHORS van Doorn, L.J., Kleter, B., Stuyver, L., Maertens, G., Brouwer, H., Schalm, S.W., Heijtsink, R. and Quint, W.
TITLE Analysis of hepatitis C virus genotypes by a line probe assay and correlation with antibody profiles
JOURNAL J. Hepatol. 21 (1), 122-129 (1994)
MEDLINE 95052487
PUBMED 7525693
AUTHORS van Doorn, L.J., Kleter, G.E., Stuyver, L., Maertens, G., Brouwer, J.T., Schalm, S.W., Heijtsink, R.A. and Quint, W.G.
TITLE Sequence analysis of hepatitis C virus genotypes 1 to 5 reveals multiple novel subtypes in the Benelux countries
JOURNAL J. Gen. Virol. 76 (Pt 7), 1871-1876 (1995)
MEDLINE 97201609
PUBMED 9049395
FEATURES
source
1. .447
/organism="Hepatitis C virus type 3"
/mol_type="genomic RNA"
/db_xref="taxon:40363"
/clone="NL96"
/note="genotype: 3"
<1. .>447
/codon_start=1
/product="precursor protein"
/protein_id="AAA67840.1"
/db_xref="GI:845498"
/translation="DGINFATGNLPGCSFSIFLALFSCLLTPTAGLEYNASGLYMW
TDCSNGSIVVEAGDIIHLPGVCVPSGSGTSCWIPVSPVAVKSCAAATSLRTH
VDMVGAATLCSALYVGLGALFLVGQGFWRHRQHWTVQDCNCI"
1. .96
/product="core protein"
/note="putative"
97. .447
/product="e1 protein"
/note="putative"

CDS
mat_peptide
mat_peptide

ORIGIN
Alignment Scores:
Pred. No.: 9,266-88 Length: 447
Score: 84.00 Matches: 84
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 56.38% Indels: 0
DB: 14 Gaps: 0

US-09-851-138C-52 (1-149) x HPCCORREAL (1-447)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 1 GACGGAATTAATTCGCAACAGGGAATTTACCTGGTTGCTCTTCTCTATCTTCCTCTCG 60

QY 21 AlaLeuPheSerCysLeuLeuThrProThrAlaGlyLeuGluTyrArgAsnAlaSerGly 40
Db 61 GCTTTGTTCTCATGCTTGTCTACCCACAGCGGGCTGGAGTACCGTAATGCTCCGGA 120
```

```

Qy 41 LeuTyrMetValThrAsnAspCysSerAsnGlySerIleValTyrGluAlaGlyAspIle 60
Db 121 CTCTACATGGTAACCTAACGACTACAGTACCGGTAGTATCGTGTATGAGCGCGGATATT 180

Qy 61 IleLeuHisLeuProGlyCysValProCysValArgSerGlyAsnThrSerArgCysTrp 80
Db 181 ATCTCCACTTACCTGGCTGTGTCCCTGCGGTACGCTCTGCAATACATCAAGATGCTGG 240

Qy 81 IleProValSer 84
Db 241 ATCTCTGTGAGC 252

RESULT 4
HPVJK070A8 1584 bp RNA linear VRL 10-FEB-1999
LOCUS
DEFINITION Hepatitis C virus isolate JK070 gene for core, env, and part of
E2/NS1, partial cds.
D49752
D49752.1 GI:1197162
VERSION
KEYWORDS core, env, and part of E2/NS1.
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (sites)
AUTHORS Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tsuruta,F.,
Legmama,L.A., Miyakawa,Y. and Mayumi,M.
TITLE Hepatitis C virus variants from Jakarta, Indonesia classifiable
into novel genotypes in the second (2e and 2f), tenth (10a) and
eleventh (11a) genetic groups
J. Gen. Virol. 77 (Pt 2), 293-301 (1996)
JOURNAL
MEDLINE 96226020
PUBMED 8627233
REFERENCE 2 (bases 1 to 1584)
AUTHORS Okamoto,H.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1584)
AUTHORS Okamoto,H.
TITLE Direct Submission
Submitted (17-MAR-1995) Hiroaki Okamoto, Jichi Medical School,
Immunology Division; Minamikawachi-machi, Kawachi-gun, Tochigi
329-04, Japan (E-mail:hokamotow@jichi.ac.jp,
Tel:0285-44-2111(ex.3334), Fax:0285-44-1557)
FEATURES
source
Location/Qualifiers
1..1584
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="JK070"
/db_xref="taxon:11103"
<1..339
/citation=[2]
/evidence=not_experimental
340..1584
/citation=[2]
/codon_start=1
/evidence=not_experimental
/product="core, env, and part of E2/NS1"
/protein_id="BAA08586.1"
/db_xref="GI:1197163"
translation="MSTLPKPKQTKNTNRPPQDKFPFGGRQIVGGVYVLPFRGPRRL
GVRAVHTKTSRQPSRRRQPIPRARTTGRSWAQPGVPWPLYGNEGCGWAGLLSPRG
SRPAGPNDPRRRNLGKVIDTTCGPAFLMGVYIPLVAPVGGVARALAGVRALED
GINATGNLPCSFSLFLLALLSCLLTPTAGLEYNVSGLYIVTNDGNSNIVYEAGD
IILHLPGVPCVRSNTRSCRTWPSPTVAVSRPGAVTASLRTHDMVNGAATLCSALY
VGDLCGALFLVGQGFQSWRHRQHTVQECNCSIIYPGLHTGHRMADWMNNSPVTMVV
SQVRLLPQTIFDLVAGRWAGVWAGVAYYSMQGNWAKVFLVLCFLPSGVDASTTISGGA
ARSVWGITSLSFSPGNSQNQLVN"

ORIGIN
Alignment Scores:
Pred. No.: 3.69e-23 Length: 1584
Score: 29.00 Matches: 29
Percent Similarity: 100.00% Conservative: 0

5' UTR
CDS
mat_peptide
mat_peptide

```

```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.46% Indels: 0
DB: 14 Gaps: 0

US-09-851-138c-52 (1-149) x HPVJK070A8 (1-1584)

Qy 52 SerIleValTyrGluAlaGlyAspIleIleLeuHisLeuProGlyCysValProCysVal 71
Db 970 AGTATTGTATGAGCGCGGATATTATCTCCACTTGCCTGGCTGTGTCCCTCGCGTA 1029

Qy 72 ArgSerGlyAsnThrSerArgCysTrp 80
Db 1030 CGCTCTGCAATACATCAAGATGTGG 1056

RESULT 5
HPCCOREE1R
LOCUS
DEFINITION Hepatitis C virus type 3 clone NL20 precursor protein gene, partial
cds.
ACCESSION L39297
VERSION L39297.1 GI:845457
KEYWORDS
SOURCE Hepatitis C virus type 3
ORGANISM Hepatitis C virus type 3
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 447)
AUTHORS van Doorn,L.J., Kleter,B., Stuyver,L., Maertens,G., Brouwer,H.,
Schalm,S., Heijtkink,R. and Quint,W.
TITLE Analysis of hepatitis C virus genotypes by a line probe assay and
correlation with antibody profiles
J. Hepatol. 21 (1), 122-129 (1994)
JOURNAL
MEDLINE 95052487
PUBMED 7525693
REFERENCE 2 (bases 1 to 447)
AUTHORS van Doorn,L.J., Kleter,G.E., Stuyver,L., Maertens,G., Brouwer,J.T.,
Schalm,S.W., Heijtkink,R.A. and Quint,W.G.
TITLE Sequence analysis of hepatitis C virus genotypes 1 to 5 reveals
multiple novel subtypes in the Benelux countries
J. Gen. Virol. 76 (Pt 7), 1871-1876 (1995)
JOURNAL
MEDLINE 97201609
PUBMED 9049395
FEATURES
source
Location/Qualifiers
1..447
/organism="Hepatitis C virus type 3"
/mol_type="genomic RNA"
/db_xref="taxon:40363"
/clones="NL20"
/notes="genotype: 3"
<1..447
/codon_start=1
/product="precursor protein"
/protein_id="AAA67820.1"
/db_xref="GI:845458"
translation="DGINFATGNLPGGSFSLFLALFSLIHPAASLEWRNTSGLYVL
TNDLNSIVVEADVDVDPGCIPCVQDGNSTCTWXTVPTVAVRYVGATTASIRSH
VDLLVGAATWCSALYVGMCGAVFLVAQAFTFRPRRHQTQVTCNCSL"
1..96
/product="core protein"
/notes="putative"
97..447
/product="e1 protein"
/notes="putative"

ORIGIN
Alignment Scores:
Pred. No.: 3.18e-20 Length: 447
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 14 Gaps: 0

```

US-09-851-138C-52 (1-149) x HPCOREE1R (1-447)		
QY	1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20 	
Db	1 GACGGGATAAATTTGCAACAGGGAACCTTGCCCGGTGCTCTCTTTCTATCTTCTCTT 60 	
QY	21 AlaLeuPheSerCysLeu 26 	
Db	61 GCTCTGTTCTCTTGCTTA 78 	
RESULT 6		
LOCUS	HPCOREE1S	447 bp ss-RNA linear VRL 16-OCT-2001
DEFINITION	Hepatitis C virus type 3a clone NL26 precursor protein gene, partial cds.	
ACCESSION	L39298	
VERSION	L39298.1	GI:845459
KEYWORDS		
SOURCE	Hepatitis C virus type 3a	
ORGANISM	Hepatitis C virus type 3a	
REFERENCE	1 (bases 1 to 447)	
AUTHORS	van Doorn,L.J., Kleter,B., Stuyver,L., Maertens,G., Brouwer,H., Schalm,S., Heijtkink,R. and Quint,W.	
TITLE	Analysis of hepatitis C virus genotypes by a line probe assay and correlation with antibody profiles	
JOURNAL	J. Hepatol. 21 (1), 122-129 (1994)	
MEDLINE	95052487	
PUBMED	7525693	
REFERENCE	2 (bases 1 to 447)	
AUTHORS	van Doorn,L.J., Kleter,G.E., Stuyver,L., Maertens,G., Brouwer,J.T., Schalm,S.W., Heijtkink,R.A. and Quint,W.G.	
TITLE	Sequence analysis of hepatitis C virus genotypes 1 to 5 reveals multiple novel subtypes in the Benelux countries	
JOURNAL	J. Gen. Virol. 76 (Pt 7), 1871-1876 (1995)	
MEDLINE	97201609	
PUBMED	9049395	
FEATURES	Location/Qualifiers	
source	1..447	
	/organism="Hepatitis C virus type 3a"	
	/mol_type="genomic RNA"	
	/db_xref="taxon:31652"	
	/clone="NL26"	
	/note="genotype: 3a"	
	<1..>447	
	/codon_start=1	
	/product="precursor protein"	
	/protein_id="AAA67821.1"	
	/db_xref="GI:845460"	
	/translation="DGINFATGNLPGCSFSIFLLALFSLIHPAASLEWRNTSGLYIL TNDCSNSIVYEADDDVILHTPGCIPCVDGNTSTCWTPTVPTVAVKYVGATTASIRSH VDLLVGQATMCSALVGMCGAVFLVQAFTFRPRRHQTVQTCNCSL"	
	1..96	
mat_peptide	/product="core protein"	
	/note="putative"	
97..447	/product="e1 protein"	
mat_peptide	/note="putative"	
ORIGIN		
Alignment Scores:		
Pred. No.:	3,18e-20	Length: 447
Score:	26.00	Matches: 26
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	17.45%	Indels: 0
DB:	14	Gaps: 0
US-09-851-138C-52 (1-149) x HPCOREE1S (1-447)		
QY	1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20 	
Db	1 GACGGGATAAATTTGCAACAGGGAACCTTGCCCGGTGCTCTCTTTCTATCTTCTTCTT 60 	
QY	21 AlaLeuPheSerCysLeu 26 	

THERAPEUTIC AND DIAGNOSTIC AGENTS
 Patent: WO 9425601-A 17 10-NOV-1994;
 INNOGENETICS NV (BE)
 Other publication CA 2139100 941110
 Other publication AU 6722294 941121
 Other publication CN 1108030 950906
 Other publication FI 946066 941223
 Other publication NO 944967 941221
 Other publication JP 7508423T 950921.
 Location/Qualifiers
 1..541
 /organism="unidentified"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32644"
 /clone="HD10-2-21"
 CDS
 2..541
 /note="unnamed protein product; protein sequence is in
 conflict with the conceptual translation"
 /codon_start=1
 /protein_id="CAA03495.1"
 /db_xref="GI:2296653"
 /translation="VGAPVGGVARALAHGVRALDEGINPATNLPGCSPSIFLLALFLS
 CLIHPAASLEWNTSGLYVLVNDSCNSSIVYEADDVILHTPGCPVCQDGNFSACWTP
 VTPVAVRYVGGATVTSIRRHVDLLVGAATMCSALYVGDMGCAVFLVGQATFRPRRH
 TVQTNCNGLYPGHLGSHRNA"
 ORIGIN
 Alignment Scores:
 Pred. No.: 3 86e-20 Length: 541
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.45% Indels: 0
 DB: 6 Gaps: 0
 US-09-851-138C-52 (1-149) x A40617 (1-541)
 QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
 |||
 Db 62 GACGGGATAAAATTCGACACGGGAATTTGCCCGTTGCTCCCTTTCTATCTCTCTT 121
 |||
 QY 21 AlaLeuPheSerCysLeu 26
 |||
 Db 122 GCTCTGTTCTCTTGCTTA 139
 |||
 RESULT 10
 A40619 541 bp DNA linear PAT 05-MAR-1997
 LOCUS
 DEFINITION
 Sequence 19 from Patent WO9425601.
 A40619
 ACCESSION
 VERSION
 A40619.1 GI:2296654
 KEYWORDS
 SOURCE
 unidentified
 unidentified
 unclassified.
 1 (bases 1 to 541)
 REFERENCE
 Maertens,G. and Stuyver,L.
 AUTHORS
 NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS
 TITLE
 THERAPEUTIC AND DIAGNOSTIC AGENTS
 JOURNAL
 Patent: WO 9425601-A 19 10-NOV-1994;
 INNOGENETICS NV (BE)
 Other publication CA 2139100 941110
 Other publication AU 6722294 941121
 Other publication CN 1108030 950906
 Other publication FI 946066 941223
 Other publication NO 944967 941221
 Other publication JP 7508423T 950921.
 Location/Qualifiers
 1..541
 /organism="unidentified"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32644"
 /clone="BR36-9-13"
 FEATURES
 source

```
CDS
2..541
/note="unnamed protein product; Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/protein_id="CAA02496.1"
/db_xref="GI:2296655"
/translation="VGAPVGGVARALAHGVRALEDGINFATGNLPGCSFSIFLLALFS
CLIHPAASLEWRNTSGLYLTNDSCNSSIVYEADDDVILHPCICVQDGNSTCWTP
VPTPVAVRYVGATTASIRSHVDLLVGAATMCSALYVGMCGAVFLVGQAFTRPRHQ
TVQTCNCSLYPGHLSGRMA"

ORIGIN
Alignment Scores:
Pred. No.: 3,86e-20 Length: 541
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-52 (1-149) x A40619 (1-541)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
DB 62 GACGGGATAAATTCGCAACAGGGAATTTGCCGGTTCCTCTTTCTATTCTCTTCTT 121
QY 21 AlaLeuPheSerCysLeu 26
DB 122 GCTCTGTTCTCTTGCTTA 139

RESULT 11
A40621
LOCUS A40621 541 bp DNA linear PAT 05-MAR-1997
DEFINITION Sequence 21 from Patent WO9425601.
ACCESSION A40621
VERSION A40621.1 GI:2296656
KEYWORDS unidentified
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 541)
AUTHORS Maertens,G. and Stuyver,L.
TITLE NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS
THERAPEUTIC AND DIAGNOSTIC AGENTS
JOURNAL Patent: WO 9425601-A 21 10-NOV-1994;
INNOGENETICS NV (BE)
COMMENT Other publication CA 2139100 941110
Other publication AU 6722294 941121
Other publication CN 1108030 950906
Other publication FI 946066 941223
Other publication NO 944967 941221
Other publication JP 7508423T 950921.
FEATURES
source
1..541
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
/clone="BR36-9-20"

CDS
2..541
/note="unnamed protein product; Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/protein_id="CAA02497.1"
/db_xref="GI:2296657"
/translation="VGAPVGGVARALAHGVRALEDGINFATGNLPGCSFSIFLLALFS
CLIHPAASLEWRNTSGLYLTNDSCNSSIVYEADDDVILHPCICVQDGNSTCWTP
VPTPVAVRYVGATTASIRSHVDLLVGAATMCSALYVGMCGAVFLVGQAFTRPRHQ
TVQTCNCSLYPGHLSGRMA"

ORIGIN
Alignment Scores:
Pred. No.: 3,86e-20 Length: 541
Score: 26.00 Matches: 26
```

```
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-52 (1-149) x A40621 (1-541)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
DB 62 GACGGGATAAATTCGCAACAGGGAATTTGCCGGTTCCTCTTTCTATTCTCTTCTT 121
QY 21 AlaLeuPheSerCysLeu 26
DB 122 GCTCTGTTCTCTTGCTTA 139

RESULT 12
A40623
LOCUS A40623 541 bp DNA linear PAT 05-MAR-1997
DEFINITION Sequence 23 from Patent WO9425601.
ACCESSION A40623
VERSION A40623.1 GI:2296658
KEYWORDS unidentified
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 541)
AUTHORS Maertens,G. and Stuyver,L.
TITLE NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS
THERAPEUTIC AND DIAGNOSTIC AGENTS
JOURNAL Patent: WO 9425601-A 23 10-NOV-1994;
INNOGENETICS NV (BE)
COMMENT Other publication CA 2139100 941110
Other publication AU 6722294 941121
Other publication CN 1108030 950906
Other publication FI 946066 941223
Other publication NO 944967 941221
Other publication JP 7508423T 950921.
FEATURES
source
1..541
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
/clone="BR33-1-10"

CDS
2..541
/note="unnamed protein product; Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/protein_id="CAA02498.1"
/db_xref="GI:2296659"
/translation="VGAPVGGVARALAHGVRALEDGINFATGNLPGCSFSIFLLALFS
CLIHPAASLEWRNTSGLYLTNDSCNSSIVYEADDDVILHPCICVQDGNSTCWTP
VPTPVAVRYVGATTASIRSHVDLLVGAATMCSALYVGMCGAVFLVGQAFTRPRHQ
TVQTCNCSLYPGHLSGRMA"

ORIGIN
Alignment Scores:
Pred. No.: 3,86e-20 Length: 541
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-52 (1-149) x A40623 (1-541)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
DB 62 GACGGGATAAATTCGCAACAGGGAATTTGCCGGTTCCTCTTTCTATTCTCTTCTT 121
QY 21 AlaLeuPheSerCysLeu 26
DB 122 GCTCTGTTCTCTTGCTTA 139
```

RESULT 13	A40625	Sequence 25 from Patent WO9425601.	541 bp	DNA	linear	PAT 05-MAR-1997
LOCUS	A40625	Sequence 25 from Patent WO9425601.	541 bp	DNA	linear	PAT 05-MAR-1997
DEFINITION	A40625	Sequence 25 from Patent WO9425601.	541 bp	DNA	linear	PAT 05-MAR-1997
ACCESSION	A40625	Sequence 25 from Patent WO9425601.	541 bp	DNA	linear	PAT 05-MAR-1997
VERSION	A40625.1	GI:2296660				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						
FEATURES						
source						
CDS						
ORIGIN						
Alignment Scores:						
Pred. No.:						
Score:						
Percent Similarity:						
Best Local Similarity:						
Query Match:						
DB:						
US-09-851-138C-52 (1-149) x A40625 (1-541)						
Qy	1	AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20				
Db	62	GACGGGATAAATTCGCAACAGGAATTGCCCGGTGCTCTTTCTATCTCTCTT 121				
Qy	21	AlaLeuPheSerCysLeu 26				
Db	122	GCTCTGTTCTTGCTTA 139				
RESULT 14	A40627	Sequence 27 from Patent WO9425601.	541 bp	DNA	linear	PAT 05-MAR-1997
LOCUS	A40627	Sequence 27 from Patent WO9425601.	541 bp	DNA	linear	PAT 05-MAR-1997
DEFINITION	A40627	Sequence 27 from Patent WO9425601.	541 bp	DNA	linear	PAT 05-MAR-1997
ACCESSION	A40627	Sequence 27 from Patent WO9425601.	541 bp	DNA	linear	PAT 05-MAR-1997
VERSION	A40627.1	GI:2296662				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						
FEATURES						
source						
CDS						
ORIGIN						
Alignment Scores:						
Pred. No.:						
Score:						
Percent Similarity:						
Best Local Similarity:						
Query Match:						
DB:						
US-09-851-138C-52 (1-149) x A40625 (1-541)						
Qy	1	AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20				
Db	62	GACGGGATAAATTCGCAACAGGAATTGCCCGGTGCTCTTTCTATCTCTCTT 121				
Qy	21	AlaLeuPheSerCysLeu 26				
Db	122	GCTCTGTTCTTGCTTA 139				

CC New sequences of hepatitis C virus genotypes for diagnosis,

CC and therapy prophylaxis

FT Key Location/Qualifiers

FT CDS 2..541.

Location/Qualifiers

1..541

/organism="unidentified"

/mol_type="genomic DNA"

/db_xref="taxon:32644"

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.:	3.86e-20	Length:	541
Score:	26.00	Matches:	26
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	17.45%	Indels:	0
DB:	6	Gaps:	0

US-09-851-138C-52 (1-149) x BD172130 (1-541)

Qy 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeu 20

Db 62 GACGGATAAATTCGCAACAGGGAATTGCCGGTTGCTCTTCTATCTTCCTTCIT 121

Qy 21 AlaLeuPheSerCysLeu 26

Db 122 GCTCTGTCTCTTGCTTA 139

Search completed: March 3, 2005, 18:33:05

Job time : 5831.52 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 14:30:42 ; Search time 1062.87 Seconds

(without alignments)
829.870 Million cell updates/sec

Title: US-09-851-138c-52

Perfect score: 149

Sequence: 1 DGINPATNLPGCSFSIFLL.....QGFSWRHRQHWTVCNCISI 149

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8776198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlp
-O=/cpn2.1/USPTO.spool_p/US09851138/runat_28022005_120306_21457/app.query.fasta_1.1123
-DB=N_Geneseq_16Dec04 -QFMT=fastap -SUFFIX=olig.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09851138 @CGN 1.1.1418 @runat_28022005_120306_21457 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04.*

- 1: Geneseqn1980a.*
- 2: Geneseqn1990a.*
- 3: Geneseqn2000a.*
- 4: Geneseqn2001a.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002a.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003a.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004a.*
- 13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84	56.4	447	2	AAT27962 Hepatitis
2	26	17.4	541	2	AQ78031 Hepatitis
3	26	17.4	541	2	AQ78033 Hepatitis
4	26	17.4	541	2	AQ78036 Hepatitis
5	26	17.4	541	2	AQ78034 Hepatitis

6	26	17.4	541	2	AAQ78035	Hepatitis
7	26	17.4	541	2	AAQ78032	Hepatitis
8	26	17.4	541	2	AAQ78029	Hepatitis
9	26	17.4	573	2	AAT16642	Hepatitis
10	26	17.4	573	2	AAT16645	Hepatitis
11	26	17.4	573	2	AAT16643	Hepatitis
12	26	17.4	573	2	AAT16644	Hepatitis
13	26	17.4	630	2	AAT12965	Hepatitis
14	26	17.4	630	6	AA148929	Hepatitis
15	26	17.4	630	10	ADD55537	Hepatitis
16	26	17.4	630	12	ADP71119	Hepatitis
17	22	14.8	447	2	AAT27951	Hepatitis
18	21	14.1	573	2	AAT16649	Hepatitis
19	21	14.1	9444	2	AAT13279	Hepatitis
20	19	12.8	541	2	AAQ78030	Hepatitis
21	19	12.8	2551	2	AAQ29630	Hepatitis
22	19	12.8	2551	2	AAQ43891	Hepatitis
23	19	12.8	9589	2	AAQ38218	Hepatitis
24	19	12.8	9611	5	AAQ86646	Hepatitis
25	19	12.8	9611	5	AAQ86645	Hepatitis
26	19	12.8	9611	5	AAQ86647	Hepatitis
27	19	12.8	9611	5	AAQ86648	Hepatitis
28	19	12.8	9711	4	AAQ23486	Hepatitis
29	19	12.8	9711	4	AAQ86937	Hepatitis
30	19	12.8	9711	5	AAQ86644	Hepatitis
31	17	11.4	90	6	ABK36447	Hepatitis
32	17	11.4	356	2	AAT27949	Hepatitis
33	17	11.4	405	2	AAQ35079	Hepatitis
34	17	11.4	407	3	AAQ75290	Hepatitis
35	17	11.4	407	12	ADN35965	Hepatitis
36	17	11.4	447	2	AAT27953	Hepatitis
37	17	11.4	447	2	AAT27955	Hepatitis
38	17	11.4	447	2	AAT27946	Hepatitis
39	17	11.4	447	2	AAT27948	Hepatitis
40	17	11.4	447	2	AAT27952	Hepatitis
41	17	11.4	447	2	AAT27938	Hepatitis
42	17	11.4	549	2	AAQ31085	Hepatitis
43	17	11.4	549	2	AAQ31089	Hepatitis
44	17	11.4	549	2	AAQ31090	Hepatitis
45	17	11.4	549	2	AAQ31094	Hepatitis

ALIGNMENTS

RESULT 1
AAT27962
ID AAT27962 standard; DNA; 447 BP.
XX
AC AAT27962;
XX
DT 11-MAR-1997 (first entry)
XX
DE Hepatitis C virus type 10a isolate NN98 bases 478-925.
XX
KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
PCR; primer; probe; antibody; infection; ss.
XX
OS Hepatitis C virus.
XX
PN WO9613590-A2.
XX
PD 09-MAY-1996.
XX
PF 23-OCT-1995; 95WO-EP004155.
XX
PR 21-OCT-1994; 94EP-00870166.
XX
PR 28-JUN-1995; 95EP-00870076.
XX
(INNO-) INNOGENETICS NV.
PA Aat27962 Hepatitis
XX Aaq78031 Hepatitis
PI Aaq78033 Hepatitis
XX Aaq78036 Hepatitis
DR WPI; 1996-251460/25.

DR P-PSDB; AAR96551.
XX Hepatitis C virus poly:nucleic acid unique to unidentified sub:type -
PT used to develop probes and primers for new sub:types and vaccines to
PT prevent and treat infection.
XX Claim 6; Fig 3; 150pp; English.
XX The sequences AAT27937-T27989 represent novel sequences isolated from
CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f, 4a-
CC 1j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l, 3g,
CC 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the genome.
CC This sequence represents nucleotides 478-925 from the HCV type 10a
CC isolate NE98. The new HCV types were isolated from patients with chronic
CC HCV from the Benelux countries, France, Cameroon and Vietnam, because of
CC their aberrant reactivities. The RNA was extracted, cDNA synthesised and
CC PCR amplified, cloned and genotyped. The 5'-UR, Core/E1 and NS5B regions
CC were sequenced either directly or partially and used to classify the new
CC viruses into (sub)types based on comparison with known sequences. The
CC sequences were used to generate the peptides AAR96424-R96524. The
CC sequences can also be used to synthesise probes and primers for the
CC detection of HCV in a sample. The polypeptides can be used to detect anti
CC -HCV antibodies, for HCV typing or to prevent HCV infections
XX
SQ Sequence 447 BP; 82 A; 130 C; 114 G; 118 T; 0 U; 3 Other;

Alignment Scores:
Pred. No.: 6,25e-76 Length: 447
Score: 84.00 Matches: 84
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 56.38% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x AAT27962 (1-447)
QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
DB 1 GACGGAAATTAATTCGCAACAGGAAATTTACCTGGTTGCTTCTCTATCTTCCTTCG 60
QY 21 AlaLeuPheSerCysLeuLeuThrProThrAlaGlyLeuGluTyArgAsnAlaSerGly 40
DB 61 GCTTTGTTCTCATGCTGTGCTTACACCCAGCGGGCTGGAGTACGTAATGCCTCCGA 120
QY 41 LeuTyMetValThrAsnAspCysSerAsnGlySerIleValTyGluAlaGlyAspIle 60
DB 121 CTCATCATGGTAACCTAACGACTGCAGTACGGTAGTAGTATCGTATGAGCGCGGGATATT 180
QY 61 IleLeuHisLeuProGlyCysValProCysValArgSerGlyAsnThrSerArgCysTrp 80
DB 181 ATCCTCCACTTACCTGGCTGTGTCCTCGCTACGCTCTGGCAATACATCAAGATGCTGG 240
QY 81 IleProValSer 84
DB 241 ATCCCTGTGAGC 252
ID AAQ78031 standard; cDNA; 540 BP.
XX AAQ78031;
XX
XX 25-MAR-2003 (revised)
DT 21-JUL-1995 (first entry)
XX Hepatitis C virus Core/E1 region.
XX Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW classification; immunisation; prophylaxis; serotyping; ss.
XX
XX Hepatitis C virus type 3a.
XX

FH Key Location/Qualifiers
FT CDS 2..541
FT /*tag= a
FT /product= "Core/E1 polypeptide."
XX
PN WO9425601-A2.
XX
PD 10-NOV-1994.
XX
XX 27-APR-1994; 94WO-EP001323.
XX
XX 27-APR-1993; 93EP-00401099.
PR 05-AUG-1993; 93EP-00402019.
XX
XX (INNO-) INNOGENETICS NV SA.
XX
XX Maertens G, Stuyver L;
XX WPI; 1994-358277/44.
DR P-PSDB; AAR63279.
XX
XX New polynucleotide sequences from hepatitis C virus - and related
PT vectors, polypeptide(s) and antibodies, useful for immunisation,
PT treatment, diagnosis and typing of HCV isolates.
XX
PS Claim 2; Page 107-108; 404pp; English.
XX
XX Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the Core/E1
CC region of HCV subtype 3a; (ii) the region spanning positions 4664-4730 of
CC the NS3 region of HCV type 3; (iii) the region spanning positions 4892-
CC 5292 of the NS3/4 region of HCV type 3; (iv) the region spanning
CC positions 8023-8235 of the NS5 region of the BR36 subgroup of HCV subtype
CC 3a; or (v) an HCV subtype 3c genomic sequence, may be used as primers to
CC amplify nucleic acid from an isolate belonging to a specific genotype, or
CC as a probe for specific detection/classification of nucleic acid.
CC Polypeptides encoded by the nucleotides in such compositions may be used
CC for immunisation against HCV, for the detection of antibodies directed
CC against HCV and for serotyping. This sequence corresponds to the Core/E1
CC region of HCV subtype 3a and is taken from a clone designated HD10-2-21.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 540 BP; 105 A; 153 C; 144 G; 138 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 7,38e-17 Length: 540
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x AAQ78031 (1-540)
QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
DB 61 GACGGGATAAATTCGCAACAGGAAATTTGCCCGGTTCCTCTTTCTATCTTCCTTCTT 120
QY 21 AlaLeuPheSerCysLeu 26
DB 121 GCTCTGTTCTCTTGCTTA 138
RESULT 3
AAQ78033
ID AAQ78033 standard; cDNA; 541 BP.
XX AAQ78033;
XX
XX 25-MAR-2003 (revised)
DT 01-AUG-1995 (first entry)
XX
XX Hepatitis C virus Core/E1 region.
DE

XX Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW classification; immunisation; prophylaxis; serotyping; ss.
XX Hepatitis C virus type 3a.

XX Key Location/Qualifiers
FH CDS 2..541
FT /*tag= a
FT /product= "Core/E1 polypeptide."

XX WO9425501-A2.

XX 10-NOV-1994.

XX 27-APR-1994; 94WO-EP001323.

XX 27-APR-1993; 93EP-00401099.

XX 05-AUG-1993; 93EP-00402019.

XX (INNO-) INNOGENETICS NV SA.

XX Maertens G, Stuyver L;

XX WPI; 1994-358277/44.

XX P-PSDB; AAR63281.

XX New polynucleotide sequences from hepatitis C virus - and related
PT vectors, polypeptide(s) and antibodies, useful for immunisation,
PT treatment, diagnosis and typing of HCV isolates.

XX Claim 2; Page 111-112; 404pp; English.

XX Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the Core/E1
CC region of HCV subtype 3a; (ii) the region spanning positions 4664-4730 of
CC the NS3 region of HCV type 3; (iii) the region spanning positions 4892-
CC 5292 of the NS3/4 region of HCV type 3; (iv) the region spanning
CC positions 8023-8235 of the NS5 region of the BR36 subgroup of HCV subtype
CC 3a; or (v) an HCV subtype 3c genomic sequence, may be used as primers to
CC amplify nucleic acid from an isolate belonging to a specific genotype, or
CC as a probe for specific detection/classification of nucleic acid.

CC Polypeptides encoded by the nucleotides in such compositions may be used
CC for immunisation against HCV, for the detection of antibodies directed
CC against HCV and for serotyping. This sequence corresponds to the Core/E1
CC region of HCV subtype 3a and is taken from a clone designated BR36-9-20.
XX (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 541 BP; 106 A; 154 C; 142 G; 139 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 7,39e-17 Length: 541
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x AAQ78033 (1-541)

QY 1 AepGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
|||||
DB 62 GACGGGATAAATTCGCACAGGGAATTTGCCGGTTCCTCTTTCTATTTCTCTCTT 121

QY 21 AlaLeuPheSerCysLeu 26
|||||

DB 122 GCTCTGTTCTCTTGCTTA 139
|||||

RESULT 4

AAQ78036

ID AAQ78036 standard; cDNA; 541 BP.

XX

AC AAQ78036;

XX 25-MAR-2003 (revised)

DT 01-AUG-1995 (first entry)

XX Hepatitis C virus Core/E1 region.

XX Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW classification; immunisation; prophylaxis; serotyping; ss.

XX Hepatitis C virus type 3a.

XX Key Location/Qualifiers
FH CDS 2..541
FT /*tag= a
FT /product= "Core/E1 polypeptide."

XX WO9425501-A2.

XX 10-NOV-1994.

XX 27-APR-1994; 94WO-EP001323.

XX 27-APR-1993; 93EP-00401099.

XX 05-AUG-1993; 93EP-00402019.

XX (INNO-) INNOGENETICS NV SA.

XX Maertens G, Stuyver L;

XX WPI; 1994-358277/44.

XX P-PSDB; AAR63284.

XX New polynucleotide sequences from hepatitis C virus - and related
PT vectors, polypeptide(s) and antibodies, useful for immunisation,
PT treatment, diagnosis and typing of HCV isolates.

XX Claim 2; Page 117-118; 404pp; English.

XX Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the Core/E1
CC region of HCV subtype 3a; (ii) the region spanning positions 4664-4730 of
CC the NS3 region of HCV type 3; (iii) the region spanning positions 4892-
CC 5292 of the NS3/4 region of HCV type 3; (iv) the region spanning
CC positions 8023-8235 of the NS5 region of the BR36 subgroup of HCV subtype
CC 3a; or (v) an HCV subtype 3c genomic sequence, may be used as primers to
CC amplify nucleic acid from an isolate belonging to a specific genotype, or
CC as a probe for specific detection/classification of nucleic acid.

CC Polypeptides encoded by the nucleotides in such compositions may be used
CC for immunisation against HCV, for the detection of antibodies directed
CC against HCV and for serotyping. This sequence corresponds to the Core/E1
CC region of HCV subtype 3a and is taken from a clone designated BR33-1-20.
XX (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 541 BP; 100 A; 154 C; 148 G; 139 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 7,39e-17 Length: 541
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x AAQ78036 (1-541)

QY 1 AepGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
|||||

DB 62 GACGGGATAAATTCGCACAGGGAATTTGCCGGTTCCTCTTTCTATTTCTCTCTT 121

QY 21 AlaLeuPheSerCysLeu 26
|||||

```

Db      122 GCTCTGTTCTCTTGCTTA 139
RESULT 5
AAQ78034
ID      AAQ78034 standard; cDNA; 541 BP.
XX      AC      AAQ78034;
XX      DT      25-MAR-2003 (revised)
XX      DT      01-AUG-1995 (first entry)
XX      DE      Hepatitis C virus Core/E1 region.
XX      KW      Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
XX      KW      classification; immunisation; prophylaxis; serotyping; ss.
XX      OS      Hepatitis C virus type 3a.
XX      FH      Key      Location/Qualifiers
XX      FT      CDS      2..541
XX      FT      /*tag= a
XX      FT      /product= "Core/E1 polypeptide."
XX      PN      WO9425601-A2.
XX      PD      10-NOV-1994.
XX      PF      27-APR-1994; 94WO-EP001323.
XX      PR      27-APR-1993; 93EP-00401099.
XX      PR      05-AUG-1993; 93EP-00402019.
XX      PA      (INNO-) INNOGENETICS NV SA.
XX      PI      Maertens G, Stuyver L;
XX      DR      WPI; 1994-358277/44.
XX      DR      P-PSDB; AAR63282.
XX      PT      New polynucleotide sequences from hepatitis C virus - and related
XX      PT      vectors, polypeptide(s) and antibodies, useful for immunisation,
XX      PS      treatment, diagnosis and typing of HCV isolates.
XX      PS      Claim 2; Page 113-114; 404pp; English.
XX      CC      Compositions comprising at least 5, and pref. 8 or more contiguous
XX      CC      nucleotides selected from an HCV type 3 genomic sequence, more
XX      CC      particularly (i) the region spanning positions 417-957 of the Core/E1
XX      CC      region of HCV subtype 3a; (ii) the region spanning positions 4664-4730 of
XX      CC      the NS3 region of HCV type 3; (iii) the region spanning positions 4892-
XX      CC      5292 of the NS3/4 region of HCV type 3; (iv) the region spanning
XX      CC      positions 8023-8235 of the NS5 region of the BR36 subgroup of HCV subtype
XX      CC      3a; or (v) an HCV subtype 3c genomic sequence, may be used as primers to
XX      CC      amplify nucleic acid from an isolate belonging to a specific genotype, or
XX      CC      as a probe for specific detection/classification of nucleic acid.
XX      CC      Polypeptides encoded by the nucleotides in such compositions may be used
XX      CC      for immunisation against HCV, for the detection of antibodies directed
XX      CC      against HCV and for serotyping. This sequence corresponds to the Core/E1
XX      CC      region of HCV subtype 3a and is taken from a clone designated BR33-1-10.
XX      CC      (Updated on 25-MAR-2003 to correct PN field.)
XX      SQ      Sequence 541 BP; 100 A; 157 C; 148 G; 136 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      7.39e-17      Length:      541
Score:          26.00      Matches:      26
Percent Similarity: 100.00%      Mismatches: 0
Best Local Similarity: 100.00%      Indels: 0
Query Match:    17.45%      Gaps: 0
DB:            2

US-09-851-138c-52 (1-149) x AAQ78034 (1-541)

1 AspGlyIleAenPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
62 GACGGGATAAACTTCGCAACAGGGAATTTGCCGGTTCCTCTTCTATCTTCTTCTT 121
122 GCTCTGTTCTCTTGCTTA 139

21 AlaLeuPheSerCysLeu 26
122 GCTCTGTTCTCTTGCTTA 139

RESULT 6
AAQ78035
ID      AAQ78035 standard; cDNA; 541 BP.
XX      AC      AAQ78035;
XX      DT      25-MAR-2003 (revised)
XX      DT      01-AUG-1995 (first entry)
XX      DE      Hepatitis C virus Core/E1 region.
XX      KW      Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
XX      KW      classification; immunisation; prophylaxis; serotyping; ss.
XX      OS      Hepatitis C virus type 3a.
XX      FH      Key      Location/Qualifiers
XX      FT      CDS      2..541
XX      FT      /*tag= a
XX      FT      /product= "Core/E1 polypeptide."
XX      PN      WO9425601-A2.
XX      PD      10-NOV-1994.
XX      PF      27-APR-1994; 94WO-EP001323.
XX      PR      27-APR-1993; 93EP-00401099.
XX      PR      05-AUG-1993; 93EP-00402019.
XX      PA      (INNO-) INNOGENETICS NV SA.
XX      PI      Maertens G, Stuyver L;
XX      DR      WPI; 1994-358277/44.
XX      DR      P-PSDB; AAR63283.
XX      PT      New polynucleotide sequences from hepatitis C virus - and related
XX      PT      vectors, polypeptide(s) and antibodies, useful for immunisation,
XX      PS      treatment, diagnosis and typing of HCV isolates.
XX      PS      Claim 2; Page 115-116; 404pp; English.
XX      CC      Compositions comprising at least 5, and pref. 8 or more contiguous
XX      CC      nucleotides selected from an HCV type 3 genomic sequence, more
XX      CC      particularly (i) the region spanning positions 417-957 of the Core/E1
XX      CC      region of HCV subtype 3a; (ii) the region spanning positions 4664-4730 of
XX      CC      the NS3 region of HCV type 3; (iii) the region spanning positions 4892-
XX      CC      5292 of the NS3/4 region of HCV type 3; (iv) the region spanning
XX      CC      positions 8023-8235 of the NS5 region of the BR36 subgroup of HCV subtype
XX      CC      3a; or (v) an HCV subtype 3c genomic sequence, may be used as primers to
XX      CC      amplify nucleic acid from an isolate belonging to a specific genotype, or
XX      CC      as a probe for specific detection/classification of nucleic acid.
XX      CC      Polypeptides encoded by the nucleotides in such compositions may be used
XX      CC      for immunisation against HCV, for the detection of antibodies directed
XX      CC      against HCV and for serotyping. This sequence corresponds to the Core/E1
XX      CC      region of HCV subtype 3a and is taken from a clone designated BR33-1-19.
XX      CC      (Updated on 25-MAR-2003 to correct PN field.)
XX      SQ      Sequence 541 BP; 100 A; 155 C; 148 G; 138 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      7.39e-17      Length:      541
Score:          26.00      Matches:      26
Percent Similarity: 100.00%      Mismatches: 0
Percent Similarity: 100.00%      Indels: 0
Percent Similarity: 100.00%      Gaps: 0

```


Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.45% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x AAQ78035 (1-541)

QY 1 AspGlyIleAsnPhaAlaThrGlyAsnLeuProGlyCySerPheSerIlePheLeuLeu 20
 DB 62 GACGGGATAAACTTCGCAACAGGGAATTTGCCGGTTCCTCTTTTCTATCTCTCTT 121

QY 21 AlaLeuPheSerCysLeu 26
 DB 122 GCTCTGTTCTCTTGCTTA 139

RESULT 7

AAQ78032
 ID AAQ78032 standard; cDNA; 541 BP.

XX AC AAQ78032;

DT 25-MAR-2003 (revised)

DT 01-AUG-1995 (first entry)

XX DE Hepatitis C virus Core/E1 region.

KW Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
 XX KW classification; immunisation; prophylaxis; serotyping; ss.

OS Hepatitis C virus type 3a.

XX FH Key Location/Qualifiers
 FT CDS 2..541
 FT /*tag= a
 FT /product= "Core/E1 polypeptide."

XX PN WO9425601-A2.

XX PD 10-NOV-1994.

XX PF 27-APR-1994; 94WO-EP001323.

XX PR 27-APR-1993; 93EP-00401099.

XX PR 05-AUG-1993; 93EP-00402019.

XX PA (INNO-) INNOGENETICS NV SA.

XX PI Maertens G, Stuyver L;

XX WPI; 1994-358277/44.

XX DR P-PSDB; AAR63280.

XX PT New polynucleotide sequences from hepatitis C virus - and related
 PT vectors, polypeptide(s) and antibodies, useful for immunisation,
 PT treatment, diagnosis and typing of HCV isolates.

XX PS Claim 2; Page 109-110; 404pp; English.

XX CC Compositions comprising at least 5, and pref. 8 or more contiguous
 CC nucleotides selected from an HCV type 3 genomic sequence, more
 CC particularly (i) the region spanning positions 417-957 of the Core/E1
 CC region of HCV subtype 3a; (ii) the region spanning positions 4664-4730 of
 CC the NS3 region of HCV type 3; (iii) the region spanning positions 4892-
 CC 5292 of the NS3/4 region of HCV type 3; (iv) the region spanning
 CC positions 8023-8235 of the NS5 region of the BR36 subgroup of HCV subtype
 CC 3a; or (v) an HCV subtype 3c genomic sequence, may be used as primers to
 CC amplify nucleic acid from an isolate belonging to a specific genotype, or
 CC as a probe for specific detection/classification of nucleic acid.
 CC Polypeptides encoded by the nucleotides in such compositions may be used
 CC for immunisation against HCV, for the detection of antibodies directed
 CC against HCV and for serotyping. This sequence corresponds to the Core/E1
 CC region of HCV subtype 3a and is taken from a clone designated BR36-9-13.
 XX (Updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 541 BP; 107 A; 155 C; 142 G; 137 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 7,398-17 Length: 541
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.45% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x AAQ78032 (1-541)

QY 1 AspGlyIleAsnPhaAlaThrGlyAsnLeuProGlyCySerPheSerIlePheLeuLeu 20
 DB 62 GACGGGATAAACTTCGCAACAGGGAATTTGCCGGTTCCTCTTTTCTATCTCTCTT 121

QY 21 AlaLeuPheSerCysLeu 26
 DB 122 GCTCTGTTCTCTTGCTTA 139

RESULT 8

AAQ78029
 ID AAQ78029 standard; cDNA; 541 BP.

XX AC AAQ78029;

XX DT 25-MAR-2003 (revised)

DT 20-JUL-1995 (first entry)

XX DE Hepatitis C virus Core/E1 region.

KW Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
 XX KW classification; immunisation; prophylaxis; serotyping; ss.

OS Hepatitis C virus type 3a.

XX FH Key Location/Qualifiers
 FT CDS 2..541
 FT /*tag= a
 FT /product= "Core/E1 polypeptide."

XX PN WO9425601-A2.

XX PD 10-NOV-1994.

XX PF 27-APR-1994; 94WO-EP001323.

XX PR 27-APR-1993; 93EP-00401099.

XX PR 05-AUG-1993; 93EP-00402019.

XX PA (INNO-) INNOGENETICS NV SA.

XX PI Maertens G, Stuyver L;

XX WPI; 1994-358277/44.

XX DR P-PSDB; AAR63277.

XX PT New polynucleotide sequences from hepatitis C virus - and related
 PT vectors, polypeptide(s) and antibodies, useful for immunisation,
 PT treatment, diagnosis and typing of HCV isolates.

XX PS Claim 2; Page 103-104; 404pp; English.

XX CC Compositions comprising at least 5, and pref. 8 or more contiguous
 CC nucleotides selected from an HCV type 3 genomic sequence, more
 CC particularly (i) the region spanning positions 417-957 of the Core/E1
 CC region of HCV subtype 3a; (ii) the region spanning positions 4664-4730 of
 CC the NS3 region of HCV type 3; (iii) the region spanning positions 4892-
 CC 5292 of the NS3/4 region of HCV type 3; (iv) the region spanning
 CC positions 8023-8235 of the NS5 region of the BR36 subgroup of HCV subtype
 CC 3a; or (v) an HCV subtype 3c genomic sequence, may be used as primers to
 CC amplify nucleic acid from an isolate belonging to a specific genotype, or
 CC as a probe for specific detection/classification of nucleic acid.

CC Polypeptides encoded by the nucleotides in such compositions may be used for immunisation against HCV, for the detection of antibodies directed against HCV and for serotyping. This sequence corresponds to the Core/E1 region of HCV subtype 3a and is taken from a clone designated HD10-2-5. CC (updated on 25-MAR-2003 to correct PN field.)

XX
SQ Sequence 541 BP; 104 A; 153 C; 145 G; 139 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 7,39e-17 Length: 541
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x AAQ78029 (1-541)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
|||
Db 62 GACGGGATAAAATTTCGCAACAGGGAATTTGCCCGGTTGCTCCTTTCTATCTTCCTTCTT 121

QY 21 AlaLeuPheSerCysLeu 26
|||
Db 122 GCTCTGTTCTCTTGCTTA 139

RESULT 9

AAT16642

ID AAT16642 standard; cDNA; 573 BP.

XX AC AAT16642;

XX DT 01-OCT-1996 (first entry)

XX DE Hepatitis C virus isolate HK10 core protein gene.

XX KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

XX KW hepatitis; ss.

XX OS Hepatitis C virus.

XX FH Key Location/Qualifiers

XX CDS 1..573

XX FT /*tag= a

XX FT /product= "core protein"

XX FT /note= "does not contain stop codon"

XX PN WO9605315-A2.

XX PD 22-FEB-1996.

XX PF 15-AUG-1995; 95WO-US010398.

XX PR 15-AUG-1994; 94US-00290665.

XX PA (USSH) US SEC DEPT HEALTH.

XX PI Bukh J, Miller RH, Purcell RH;

XX DR WPI; 1996-139709/14.

XX DR P-PSDB; AAR92968.

XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to

XX determine HCV genotype and as vaccines against HCV infection.

XX PS Claim 3; Page 169; 340pp; English.

XX CC AAT16610-T16661 are cDNAs encoding a core protein gene of 52 HCV

XX isolates. The isolated sequences are useful for the prodn. of primers

XX also useful for detecting the presence of HCV in a sample, the primers are

XX in vaccines for immunising against HCV infection. The proteins may also

XX be used to detect antibodies against HCV in serum, saliva, lymphocytes or

CC other mononuclear cells. The antibodies may be used in the prevention of

CC HCV infection

XX SQ Sequence 573 BP; 111 A; 179 C; 166 G; 117 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7,81e-17 Length: 573
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x AAT16642 (1-573)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20

|||

Db 478 GACGGGATAAAATTTCGCAACAGGGAATTTGCCCGGTTGCTCCTTTCTATCTTCCTTCTT 537

QY 21 AlaLeuPheSerCysLeu 26

|||

Db 538 GCTCTGTTCTCTTGCTTA 555

RESULT 10

AAT16645

ID AAT16645 standard; cDNA; 573 BP.

XX AC AAT16645;

XX DT 01-OCT-1996 (first entry)

XX DE Hepatitis C virus isolate DK12 core protein gene.

XX KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

XX KW hepatitis; ss.

XX OS Hepatitis C virus.

XX FH Key Location/Qualifiers

XX CDS 1..573

XX FT /*tag= a

XX FT /product= "core protein"

XX FT /note= "does not contain stop codon"

XX PN WO9605315-A2.

XX PD 22-FEB-1996.

XX PF 15-AUG-1995; 95WO-US010398.

XX PR 15-AUG-1994; 94US-00290665.

XX PA (USSH) US SEC DEPT HEALTH.

XX PI Bukh J, Miller RH, Purcell RH;

XX DR WPI; 1996-139709/14.

XX DR P-PSDB; AAR92971.

XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to

XX determine HCV genotype and as vaccines against HCV infection.

XX PS Claim 3; Page 170; 340pp; English.

XX CC AAT16610-T16661 are cDNAs encoding a core protein gene of 52 HCV

XX isolates. The isolated sequences are useful for the prodn. of primers

XX also useful for detecting the presence of HCV in a sample, the primers are

XX in vaccines for immunising against HCV infection. The proteins may also

XX be used to detect antibodies against HCV in serum, saliva, lymphocytes or

XX other mononuclear cells. The antibodies may be used in the prevention of

XX HCV infection

SQ Sequence 573 BP; 108 A; 179 C; 170 G; 116 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7.81e-17 Length: 573
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.45% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x AAT16645 (1-573)

Qy 1 AspGlyIleAenPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
 Db 478 GACGGGATAAATTTGCAACAGGGAACCTGGCCGGTTGCTCTTTTCTATCTTCCTTCTT 537

Qy 21 AlaLeuPheSerCysLeu 26

Db 538 GCTCTGTTCTCTCTGCTTA 555

RESULT 11

AAT16643

ID AAT16643 standard; cDNA; 573 BP.

XX

AC AAT16643;

DT 01-OCT-1996 (first entry)

DE Hepatitis C virus isolate S52 core protein gene.

XX HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis; ss.
 OS Hepatitis C virus.

XX Key
 FH CDS

Location/Qualifiers
 1..573

/tag= a

/product= "core protein"

/note= "does not contain stop codon"

XX WO9605315-A2.

XX PD 22-FEB-1996.

XX PF 15-AUG-1995; 95WO-US010398.

XX PR 15-AUG-1994; 94US-00290665.

XX PA (USSH) US SEC DEPT HEALTH.

XX PI Bukh J, Miller RH, Purcell RH;

XX WPI; 1996-139709/14.

XX P-PSDB; AAR92969.

XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to

XX determine HCV genotype and as vaccines against HCV infection.

XX Claim 3; Page 169; 340pp; English.

XX AAT16610-T16661 are cDNAs encoding a core protein gene of 52 HCV

XX isolates. The isolated sequences are useful for the prodn. of primers

XX useful for detecting the presence of HCV in a sample, the primers are

XX also useful for HCV genotyping. Proteins encoded by the cDNAs can be used

XX in vaccines for immunising against HCV infection. The proteins may also

XX be used to detect antibodies against HCV in serum, saliva, lymphocytes or

XX other mononuclear cells. The antibodies may be used in the prevention of

XX HCV infection

SQ Sequence 573 BP; 109 A; 177 C; 168 G; 119 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7.81e-17 Length: 573
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0

Pred. No.: 7.81e-17 Length: 573

Score: 26.00 Matches: 26

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 17.45% Indels: 0

DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x AAT16643 (1-573)

Qy 1 AspGlyIleAenPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20

Db 478 GACGGGATAAATTTGCAACAGGGAACCTGGCCGGTTGCTCTTTTCTATCTTCCTTCTT 537

Qy 21 AlaLeuPheSerCysLeu 26

Db 538 GCTCTGTTCTCTGCTTA 555

RESULT 12

AAT16644

ID AAT16644 standard; cDNA; 573 BP.

XX

AC AAT16644;

DT 01-OCT-1996 (first entry)

DE Hepatitis C virus isolate S2 core protein gene.

XX HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis; ss.
 OS Hepatitis C virus.

XX Key
 FH CDS

Location/Qualifiers
 1..573

/tag= a

/product= "core protein"

/note= "does not contain stop codon"

XX WO9605315-A2.

XX PD 22-FEB-1996.

XX PF 15-AUG-1995; 95WO-US010398.

XX PR 15-AUG-1994; 94US-00290665.

XX PA (USSH) US SEC DEPT HEALTH.

XX PI Bukh J, Miller RH, Purcell RH;

XX WPI; 1996-139709/14.

XX P-PSDB; AAR92970.

XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to

XX determine HCV genotype and as vaccines against HCV infection.

XX Claim 3; Page 170; 340pp; English.

XX AAT16610-T16661 are cDNAs encoding a core protein gene of 52 HCV

XX isolates. The isolated sequences are useful for the prodn. of primers

XX useful for detecting the presence of HCV in a sample, the primers are

XX also useful for HCV genotyping. Proteins encoded by the cDNAs can be used

XX in vaccines for immunising against HCV infection. The proteins may also

XX be used to detect antibodies against HCV in serum, saliva, lymphocytes or

XX other mononuclear cells. The antibodies may be used in the prevention of

XX HCV infection

SQ Sequence 573 BP; 111 A; 178 C; 166 G; 118 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7.81e-17 Length: 573
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0

```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 2 Gaps: 0

US-09-851-138c-52 (1-149) x AAT16644 (1-573)
QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
DB 478 GACGGGATAAAATTTTCACACAGGAACTTGC CGGTGCTCTTTTCTATCTCTCTTCT 537
QY 21 AlaLeuPheSerCysLeu 26
DB 538 GCCTCTGTTCTCTTGCTTA 555

RESULT 13
AAT12965
ID AAT12965 standard; DNA; 630 BP.
XX
AC AAT12965;
XX
DT 24-SEP-1996 (first entry)
XX
DE HCV E1 construct HCC162.
XX
KW HCV; E1; E2; disulphide bond cleavage; envelope protein; vaccine; human;
KW serotype; reversed phase hybridisation assay; genotype; antigen; sera;
KW SB.
XX
OS Hepatitis C virus.
XX
PN WO9604385-A2.
XX
PD 15-FEB-1996.
XX
PF 31-JUL-1995; 95WO-EP003031.
XX
PR 29-JUL-1994; 94EP-00870132.
XX
PA (INNO-) INNOGENETICS NV.
XX
PI Maertens G, Bosman F, De Martynoff G, Buyse M;
XX WPI; 1996-129401/13.
XX
DR Purifying recombinant hepatitis C virus (HCV) E1 and E2 envelope proteins
PT - in presence of disulphide bond cleavage agent, to produce proteins
PT suitable for direct use in vaccines or diagnostic assays of HCV.
XX
PS Claim 23; Fig 21; 146pp; English.
XX
CC AAT12704-T12709 and AAT12961-T12974 represent hepatitis C virus (HCV) E1
CC and E2 protein coding sequence constructs. These sequences are included
CC in vectors for the production of recombinant E1, E2, and E1/E2 proteins.
CC The recombinant proteins can then be isolated using a method of the
CC invention. In the method, the envelope proteins are purified by carrying
CC out a disulphide bond cleavage, or a reduction step with a disulphide
CC bond cleavage agent, after lysis of recombinant host cells. The
CC constructs containing the purified HCV envelope proteins can be used for
CC vaccinating humans against HCV, for in vitro detection of HCV antibodies
CC in a sample, and in a serotyping assay for detecting one or more
CC serological types of HCV present in a biological sample. The constructs
CC can also be immobilised on a solid substrate and incorporated into a
CC reversed phase hybridisation assay for determining the presence or the
CC genotype of HCV. The new purification method preserves the conformation
CC of the recombinantly expressed E1, E2 and E1/E2, and eliminates
CC contaminating proteins. Antigens isolated using this method are more
CC reactive with human sera than those isolated by known techniques
XX
SQ Sequence 630 BP; 127 A; 175 C; 168 G; 160 T; 0 U; 0 Other;

Alignment Scores: 8.56e-17 Length: 630
Pred. No.: 26.00 Matches: 26
Score:
Percent Similarity: 100.00%
Best Local Similarity: 100.00% Mismatches: 0

Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Matches: 26
Query Match: 17.45% Indels: 0
DB: 2 Gaps: 0

US-09-851-138c-52 (1-149) x AAT12965 (1-630)
QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
DB 124 GACGGGATAAAATTTTCGCAACAGGAAATTTGCCCGGTGCTCTTTTCTATCTCTCTCTC 183
QY 21 AlaLeuPheSerCysLeu 26
DB 184 GCTCTGTTCTCTTGCTTA 201

RESULT 14
AAL48929
ID AAL48929 standard; DNA; 630 BP.
XX
AC AAL48929;
XX
DT 24-OCT-2002 (first entry)
XX
DE Hepatitis C virus clone HCC162 E1 protein coding sequence.
XX
KW Hepatitis C virus; HCV; E1 protein; E2 protein; infection; gene;
KW virucide; immunostimulant; vaccine; ds.
XX
OS Hepatitis C virus.
XX
PN WO200255548-A2.
XX
PD 18-JUL-2002.
XX
PF 11-JAN-2002; 2002WO-EP000219.
XX
PR 11-JAN-2001; 2001US-0260669P.
XX
PR 30-AUG-2001; 2001US-0315768P.
XX
PA (INNO-) INNOGENETICS NV.
XX
PI Maertens G, Bosman F, Buyse M;
XX WPI; 2002-599657/64.
XX P-PSDB; AAO18670.
XX
DR New therapeutic vaccine compositions comprising at least one purified
PT recombinant hepatitis C virus (HCV) single or specific oligomeric
PT recombinant envelope protein E1 or E2, useful for immunizing humans from
PT HCV infection.
XX
PS Example 2; Page 181-182; 243pp; English.
XX
CC The present invention relates to new therapeutic vaccine compositions for
CC inducing hepatitis C virus (HCV)-specific antibodies, comprising a
CC composition containing at least one purified recombinant HCV single or
CC specific oligomeric recombinant envelope proteins selected from an E1 and
CC an E2 protein, and optionally a pharmaceutical adjuvant. The vaccines are
CC useful for inducing HCV-specific antibodies or for immunising humans
CC against HCV. The recombinant HCV E1 and/or E2 proteins are useful as
CC vaccines or therapeutics, in HCV screening and confirmatory antibody
CC tests, for raising antibodies, in the preparation of medicament, and for
CC in vitro monitoring of HCV disease or prognosing the response to
CC treatment of patients suffering from HCV infection. The present sequence
CC is a coding sequence described in the exemplification of the invention
XX
SQ Sequence 630 BP; 127 A; 175 C; 168 G; 160 T; 0 U; 0 Other;
```

Query Match: 17.45% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-52 (1-149) x AAL48929 (1-630)

Oy 1 AspGlyIleAanPheAlaThrClyAenLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 124 GACGGATAAAATTCGCAACAGGAATTGCCCGGTGCTCTTTTCATTTCCTTC 183
Oy 21 AlaLeuPheSerCysLeu 26
Db 184 GCTCTGTTCTCTTGCTTA 201

RESULT 15

ADD55537
ID ADD55537 standard; DNA; 630 BP.

XX AC ADD55537;
XX DT 15-JAN-2004 (first entry)
XX DE Hepatitis C virus E1/E2 protein coding sequence #5.
XX KW Hepatitis C virus; HCV; vaccine; liver disease; E1 protein; E2 protein;
XX KW liver fibrosis; ds; gene.
XX OS Hepatitis C virus.
XX PN WO2003051912-A2.
XX PD 26-JUN-2003.
XX PF 18-DEC-2002; 2002WO-BP014480.
XX PR 18-DEC-2001; 2001US-00020510.
XX PR 16-OCT-2002; 2002US-0418358P.
XX PA (INNO-) INNOGENETICS NV.

XX PI Maertens G, Depla E, Bosman F;
XX DR WPI; 2003-541632/51.
XX DR P-PSDB; ADD55538.

XX PT New hepatitis C virus (HCV) vaccine composition, useful for reducing
XX PT liver disease, e.g., liver fibrosis in a chronic HCV-infected mammal.

XX PS Example 2; SEQ ID NO 29; 271pp; English.
XX CC The invention comprises an Hepatitis C virus (HCV) vaccine for reducing
XX CC liver disease. The vaccine of the invention comprises an HCV E1 or E2
XX CC protein as an antigen. The HCV vaccine is useful for reducing liver
XX CC disease (e.g. liver fibrosis) in a chronic HCV-infected mammal. The
XX CC present DNA sequence encodes an HCV E1/E2 protein.

XX SQ Sequence 630 BP; 127 A; 175 C; 168 G; 160 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8,56e-17 Length: 630
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 10 Gaps: 0

US-09-851-138C-52 (1-149) x ADD55537 (1-630)

Oy 1 AspGlyIleAanPheAlaThrGlyAenLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 124 GACGGATAAAATTCGCAACAGGAATTGCCCGGTGCTCTTTTCATTTCCTTC 183
Oy 21 AlaLeuPheSerCysLeu 26
Db 184 GCTCTGTTCTCTTGCTTA 201

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:54:32 ; Search time 307.169 Seconds

(without alignments)
793.716 Million cell updates/sec

Title: US-09-851-138c-52

Perfect score: 149

Sequence: 1 DGINPATGNLPCSCSIFLL.....QGFSWRHQHWTVCNCISI 149

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2400006

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool_p/US09851138/runat_28022005_120307_21495/app_query.fasta_1.1123
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=olig.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MAIRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USERS=US09851138 @CGN 1.1.249 @runat_28022005_120307_21495 -NCPUS=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPHLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	56.4	447	3	US-08-836-075A-51
2	26	17.4	541	4	US-09-878-281A-13
3	26	17.4	541	4	US-09-878-281A-17
4	26	17.4	541	4	US-09-878-281A-19
5	26	17.4	541	4	US-09-878-281A-21
6	26	17.4	541	4	US-09-878-281A-23
7	26	17.4	541	4	US-09-878-281A-25
8	26	17.4	541	4	US-09-878-281A-27
9	26	17.4	573	2	US-08-290-665A-135
10	26	17.4	573	2	US-08-290-665A-136
11	26	17.4	573	2	US-08-290-665A-137
12	26	17.4	573	2	US-08-290-665A-138

ALIGNMENTS

RESULT 1

US-08-836-075A-51
; Sequence 51, Application US/08836075A
; Patent No. 6180768

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; APPLICANT: STUYVER, LIEVEN

; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES

; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC

; NUMBER OF SEQUENCES: 207

; CORRESPONDENCE ADDRESS:

; ADDRESS: ARNOLD, WHITE & DURKEE

; STREET: P.O. BOX 4433

; CITY: HOUSTON

; STATE: TEXAS

; COUNTRY: USA

; ZIP: 77210-4433

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Microsoft Word 6.0 / ASCII text output

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/836,075A

; FILING DATE: 21 Apr 1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/EP95/04155

; FILING DATE: 23 Oct 1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 94870166.9

; FILING DATE: 21 Oct 1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 95870076.7

; FILING DATE: 28 Jun 1995

; ATTORNEY/AGENT INFORMATION:

NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-836-075A-51

Alignment Scores:
Pred. No.: 2,53e-75 Length: 447
Score: 84.00 Matches: 84
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 56.38% Indels: 0
Gaps: 3

US-09-851-138C-52 (1-149) x US-08-836-075A-51 (1-447)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 1 GACGGAAATAATTGCAACACAGGAATTTACCTGGTTGCTCTTCTCTATCTTCCTCTG 60
QY 21 AlaLeuPheSerCysLeuThrProThrAlaGlyLeuGluTyrrArgAsnAlaSerGly 40
Db 61 GCTTTGTTCTCATGCTTGTCTACCCACACACCGCGGTGAGTACCGTAATGCTCCGGA 120
QY 41 LeuTyrrMetValThrAsnAspCysSerAsnGlySerIleValTyrrGluAlaGlyAspIle 60
Db 121 CTCATCATGTTACTACTACGACTGCATGACGTAGTATCGTATGAGCGCGGGATATT 180
QY 61 IleLeuHisLeuProGlyCysValProCysValArgSerGlyAsnThrSerArgCysTrp 80
Db 181 ATCCTCCACTTACCTGGTGTGTCCTCCCTGCGTACGCTCTGGCAATACATCAAGATGCTGG 240
QY 81 IleProValSer 84
Db 241 ATCCCTGTGAGC 252

RESULT 2
US-09-878-281A-13
; Sequence 13, Application US/09878281A
; Patent No. 6762024
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: New sequences of hepatitis C virus genotypes for diagnosis, proph
; TITLE OF INVENTION: and therapy
; FILE REFERENCE: 35
; CURRENT APPLICATION NUMBER: US/09/878,281A
; CURRENT FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 541
; TYPE: DNA
; ORGANISM: hepatitis C virus
US-09-878-281A-13

Alignment Scores:
Pred. No.: 4,49e-17 Length: 541
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
Gaps: 4

US-09-851-138C-52 (1-149) x US-09-878-281A-13 (1-541)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 62 GACGGATAAAATTCGCAACACAGGAATTTGCCGGTTGCTCTCTTCTATCTTCCTCT 121
QY 21 AlaLeuPheSerCysLeu 26
Db 122 GCTCTGTCTCTTGTCTTA 139

RESULT 3
US-09-878-281A-17
; Sequence 17, Application US/09878281A
; Patent No. 6762024
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: New sequences of hepatitis C virus genotypes for diagnosis, proph
; TITLE OF INVENTION: and therapy
; FILE REFERENCE: 35
; CURRENT APPLICATION NUMBER: US/09/878,281A
; CURRENT FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 541
; TYPE: DNA
; ORGANISM: hepatitis C virus
US-09-878-281A-17

Alignment Scores:
Pred. No.: 4,49e-17 Length: 541
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
Gaps: 4

US-09-851-138C-52 (1-149) x US-09-878-281A-17 (1-541)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 62 GACGGATAAAATTCGCAACACAGGAATTTGCCGGTTGCTCTCTTCTATCTTCCTCT 121
QY 21 AlaLeuPheSerCysLeu 26
Db 122 GCTCTGTCTCTTGTCTTA 139

RESULT 4
US-09-878-281A-19
; Sequence 19, Application US/09878281A
; Patent No. 6762024
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: New sequences of hepatitis C virus genotypes for diagnosis, proph
; TITLE OF INVENTION: and therapy
; FILE REFERENCE: 35
; CURRENT APPLICATION NUMBER: US/09/878,281A
; CURRENT FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 541
; TYPE: DNA
; ORGANISM: hepatitis C virus
US-09-878-281A-19

Alignment Scores:
Pred. No.: 4,49e-17 Length: 541
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
Gaps: 4

US-09-851-138C-52 (1-149) x US-09-878-281A-19 (1-541)

US-09-851-138C-52 (1-149) x US-09-878-201A-27 (1-541)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
 Db 62 GACGGGATAAACTTCGCAACAGGAATTTGCCGGTTCCTTTTCTATCTTCCTCTT 121

QY 21 AlaLeuPheSerCysLeu 26
 Db 122 GCTCTGTTCTCTGCTTA 139

RESULT 9

US-08-290-665A-135
 ; Sequence 135, Application US/08290665A
 ; Patent No. 5882852
 ; GENERAL INFORMATION:
 ; APPLICANT: BUKH, J., MILLER, R.H. AND
 ; APPLICANT: PURCELL, R.H.
 ; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 ; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
 ; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 ; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 ; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
 ; NUMBER OF SEQUENCES: 263
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/290,665A
 ; FILING DATE: 15-AUG-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: RICHARD W. BORK
 ; REGISTRATION NUMBER: 36,459
 ; REFERENCE/DOCKET NUMBER: 2026-4116
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; TELEX: 421792
 ; INFORMATION FOR SEQ ID NO: 135:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 573 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; ORGANISM: homosapiens
 ; INDIVIDUAL ISOLATE: HK10
 ; US-08-290-665A-135

Alignment Scores:
 Pred. No.: 4,74e-17 Length: 573
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.45% Indels: 0
 DB: 2 Gaps: 0
 US-09-851-138C-52 (1-149) x US-08-290-665A-135 (1-573)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
 Db 478 GACGGGATAAACTTCGCAACAGGAATTTGCCGGTTCCTTTTCTATCTTCCTCTT 537

QY 21 AlaLeuPheSerCysLeu 26
 Db 538 GCTCTGTTCTCTGCTTA 555

RESULT 10

US-08-290-665A-136
 ; Sequence 136, Application US/08290665A
 ; Patent No. 5882852
 ; GENERAL INFORMATION:
 ; APPLICANT: BUKH, J., MILLER, R.H. AND
 ; APPLICANT: PURCELL, R.H.
 ; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 ; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
 ; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 ; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 ; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
 ; NUMBER OF SEQUENCES: 263
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/290,665A
 ; FILING DATE: 15-AUG-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: RICHARD W. BORK
 ; REGISTRATION NUMBER: 36,459
 ; REFERENCE/DOCKET NUMBER: 2026-4116
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; TELEX: 421792
 ; INFORMATION FOR SEQ ID NO: 136:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 573 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; ORGANISM: homosapiens
 ; INDIVIDUAL ISOLATE: S52
 ; US-08-290-665A-136

Alignment Scores:
 Pred. No.: 4,74e-17 Length: 573
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.45% Indels: 0
 DB: 2 Gaps: 0
 US-09-851-138C-52 (1-149) x US-08-290-665A-136 (1-573)

QY 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
 Db 478 GACGGGATAAACTTCGCAACAGGAATTTGCCGGTTCCTTTTCTATCTTCCTCTT 537

QY 21 AlaLeuPheSerCysLeu 26
 Db 538 GCTCTGTTCTCTGCTTA 555

RESULT 11

US-08-290-665A-137
 ; Sequence 137, Application US/08290665A

Patent No. 5882852
GENERAL INFORMATION: BUKH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CORE GENES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 137:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S2
US-08-290-665A-137

Alignment Scores:
Pred. No.: 4,74e-17 Length: 573
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x US-08-290-665A-137 (1-573)
Qy 1 AspGlylleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 478 GACGGGATAAAATTTGCAACAGGGAACTTGCCGGTTCCTCTTTTCTATCTCTCTT 537

RESULT 12
US-08-290-665A-138
Sequence 138, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CORE GENES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 137:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S2
US-08-290-665A-137

Alignment Scores:
Pred. No.: 4,74e-17 Length: 573
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x US-08-290-665A-137 (1-573)
Qy 1 AspGlylleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 478 GACGGGATAAAATTTGCAACAGGGAACTTGCCGGTTCCTCTTTTCTATCTCTCTT 537

RESULT 13
US-08-290-665A-138
Sequence 135, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CORE GENES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 138:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: DK12
US-08-290-665A-138

Alignment Scores:
Pred. No.: 4,74e-17 Length: 573
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x US-08-290-665A-138 (1-573)
Qy 1 AspGlylleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 478 GACGGGATAAAATTTGCAACAGGGAACTTGCCGGTTCCTCTTTTCTATCTCTCTT 537

Qy 21 AlaLeuPheSerCysLeu 26
Db 538 GCTCTGTCTCTTGCTTA 555

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 138:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: DK12
US-08-290-665A-138

Alignment Scores:
Pred. No.: 4,74e-17 Length: 573
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x US-08-290-665A-138 (1-573)
Qy 1 AspGlylleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 478 GACGGGATAAAATTTGCAACAGGGAACTTGCCGGTTCCTCTTTTCTATCTCTCTT 537

Qy 21 AlaLeuPheSerCysLeu 26
Db 538 GCTCTGTCTCTTGCTTA 555

RESULT 13
PCT-US95-10398-135
Sequence 135, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CORE GENES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 138:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: DK12
US-08-290-665A-138

Alignment Scores:
Pred. No.: 4,74e-17 Length: 573
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-52 (1-149) x US-08-290-665A-138 (1-573)
Qy 1 AspGlylleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 478 GACGGGATAAAATTTGCAACAGGGAACTTGCCGGTTCCTCTTTTCTATCTCTCTT 537

Qy 21 AlaLeuPheSerCysLeu 26
Db 538 GCTCTGTCTCTTGCTTA 555

```

/ COUNTRY: USA
/ ZIP: 10154
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: FLOPPY DISK
/ COMPUTER: IBM PC COMPATIBLE
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: WORDPERFECT 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/10398
/ FILING DATE: 15-AUG-1995
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/086,428
/ FILING DATE: 29 JUNE 1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/290/665
/ FILING DATE: 15 AUGUST 1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: RICHARD W. BORK
/ REGISTRATION NUMBER: 36,459
/ REFERENCE/DOCKET NUMBER: 2026-4116
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 758-4800
/ TELEFAX: (212) 751-6849
/ TELEX: 421792
/ INFORMATION FOR SEQ ID NO: 135:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 573 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ ORIGINAL SOURCE:
/ ORGANISM: homosapiens
/ INDIVIDUAL ISOLATE: HK10
/ PCT-US95-10398-135

```

```

Alignment Scores:
Pred. No.: 4.74e-17 Length: 573
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 5 Gaps: 0

```

US-09-851-138C-52 (1-149) x PCT-US95-10398-135 (1-573)

```

QY 1 AspGlyVileasnPheAlaThrGlyAsnLeuProGlyCysSerPheSerilePheLeuLeu 20
Db 478 GACGGGATAAATTTGCAACAGGAACTTGCCCGGTGCTCTTTCTATCTTCCTTCTT 537
QY 21 AlaLeuPheSerCysLeu 26
Db 538 GCTCTGTTCTCTGCTTA 555

```

```

RESULT 14
PCT-US95-10398-136
/ Sequence 136, Application PC/TUS9510398
/ GENERAL INFORMATION:
/ APPLICANT: BUKH, J., MILLER, R.H. AND
/ APPLICANT: PURCELL, R.H.
/ TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
/ TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
/ TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
/ TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
/ TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
/ NUMBER OF SEQUENCES: 263
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: MORGAN & FINNEGAN
/ STREET: 345 PARK AVENUE
/ CITY: NEW YORK
/ STATE: NEW YORK
/ COUNTRY: USA
/ ZIP: 10154

```

```

/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: FLOPPY DISK
/ COMPUTER: IBM PC COMPATIBLE
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: WORDPERFECT 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/10398
/ FILING DATE: 15-AUG-1995
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/086,428
/ FILING DATE: 29 JUNE 1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/290/665
/ FILING DATE: 15 AUGUST 1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: RICHARD W. BORK
/ REGISTRATION NUMBER: 36,459
/ REFERENCE/DOCKET NUMBER: 2026-4116
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 758-4800
/ TELEFAX: (212) 751-6849
/ TELEX: 421792
/ INFORMATION FOR SEQ ID NO: 136:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 573 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ ORIGINAL SOURCE:
/ ORGANISM: homosapiens
/ INDIVIDUAL ISOLATE: S52
/ PCT-US95-10398-136

```

```

Alignment Scores:
Pred. No.: 4.74e-17 Length: 573
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 5 Gaps: 0

```

US-09-851-138C-52 (1-149) x PCT-US95-10398-136 (1-573)

```

QY 1 AspGlyVileasnPheAlaThrGlyAsnLeuProGlyCysSerPheSerilePheLeuLeu 20
Db 478 GACGGGATAAATTTGCAACAGGAACTTGCCCGGTGCTCTTTCTATCTTCCTTCTT 537
QY 21 AlaLeuPheSerCysLeu 26
Db 538 GCTCTGTTCTCTGCTTA 555

```

```

RESULT 15
PCT-US95-10398-137
/ Sequence 137, Application PC/TUS9510398
/ GENERAL INFORMATION:
/ APPLICANT: BUKH, J., MILLER, R.H. AND
/ APPLICANT: PURCELL, R.H.
/ TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
/ TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
/ TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
/ TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
/ TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
/ NUMBER OF SEQUENCES: 263
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: MORGAN & FINNEGAN
/ STREET: 345 PARK AVENUE
/ CITY: NEW YORK
/ STATE: NEW YORK
/ COUNTRY: USA
/ ZIP: 10154
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: FLOPPY DISK

```

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 137:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homoapiens
INDIVIDUAL ISOLATE: S2
PCT-US95-10398-137

Alignment Scores:
Pred. No.: 4,748-17 Length: 573
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 17.45% Indels: 0
DB: 5 Gaps: 0

US-09-851-138C-52 (1-149) x PCT-US95-10398-137 (1-573)

Oy 1 AspGlyIleAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeu 20
Db 478 GACGGGATAAAATTTCACACAGGAACCTGCCCGGTGCTCTTTTCTATCTTCCTCTT 537
Oy 21 AlaLeuPheSerCysLeu 26
Db 538 GCCCTGTCTCTTGCTTA 555

Search completed: March 3, 2005, 22:04:58
Job time : 309.169 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:43:48 ; Search time 9389.29 Seconds
(without alignments)
604.047 Million cell updates/sec

Title: US-09-851-138C-52

Perfect score: 149

Sequence: 1 DGINFATGNLPGCSFSIFLL.....QGFWRHRQHWTVQDCNCIS 149

Scoring table: QIGAPOP

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Word size: 1

Total number of hits satisfying chosen parameters: 68477535

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool_p/US09851138/runat_28022005_120306_21476/app.query.fasta_1.1123
-DB=EST -QFMT=fasta -SUFFIX=olig.rst -MINMATCH=0.1 -LOOPC=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09851138 @CGN 1 1 10973 @runat_28022005_120306_21476 -NCPU=6 -ICPU=3
-NO.WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=1120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	10	6.7	286	7 T38854	EST104418 S
C 2	10	6.7	1135	9 CL134102	CL134102 ISB1-104H
C 3	9	6.0	181	9 CC504025	CC504025 CH240_343
C 4	9	6.0	286	6 CD143793	CD143793 MGI-0088U
C 5	9	6.0	312	8 AQ844124	AQ844124 LMAJFV1_1
C 6	9	6.0	349	8 BH258325	BH258325 CH230-60P
C 7	9	6.0	372	8 AQ845487	AQ845487 LMAJFV1_1
C 8	9	6.0	413	1 AI607477	AI607477 mr84b07.y
C 9	9	6.0	414	1 AI946531	AI946531 bs27e02.y

C	10	9	6.0	446	7	CO299961	CO299961 EK177107.
C	11	9	6.0	456	7	CO325948	CO325948 EP02947.3
C	12	9	6.0	477	7	H94357	H94357 YW54605.r1
C	13	9	6.0	484	2	BF42449	BF42449 259170 MA
C	14	9	6.0	522	9	CL369856	CL369856 RPCI144_29
C	15	9	6.0	549	5	BQ039335	BQ039335 gd09a08.y
C	16	9	6.0	620	6	CD469353	CD469353 Leuko82.3
C	17	9	6.0	670	4	BM622715	BM622715 170006874
C	18	9	6.0	688	4	BJ168506	BJ168506 BJ168506
C	19	9	6.0	699	8	BZ115200	BZ115200 CH230-510
C	20	9	6.0	707	6	CD901971	CD901971 G356.105I
C	21	9	6.0	707	8	BH968902	BH968902 od105a07.
C	22	9	6.0	726	4	BM647152	BM647152 170006873
C	23	9	6.0	745	9	AG421802	AG421802 Mus muscu
C	24	9	6.0	766	9	AG541630	AG541630 Mus muscu
C	25	9	6.0	799	8	AQ745926	AQ745926 HS 2275.A
C	26	9	6.0	816	2	AW940256	AW940256 GH07353.3
C	27	9	6.0	850	9	CC942652	CC942652 BOLDW81TF
C	28	9	6.0	874	2	BF381817	BF381817 601816069
C	29	9	6.0	888	5	BQ962268	BQ962268 AGENCOURT
C	30	9	6.0	1023	4	BG027047	BG027047 602293902
C	31	9	6.0	1260	9	AG277831	AG277831 Mus muscu
C	32	9	6.0	1956	4	BG165646	BG165646 602345142
C	33	9	6.0	3542	3	AK086405	AK086405 Mus muscu
C	34	8	5.4	89	8	BH406398	BH406398 RPCI-23-4
C	35	8	5.4	132	2	AW787453	AW787453 945008H07
C	36	8	5.4	164	8	AQ469119	AQ469119 CITBI-E1-
C	37	8	5.4	189	4	BF997327	BF997327 RC0-GN013
C	38	8	5.4	193	1	AI595178	AI595178 mk09d12.y
C	39	8	5.4	196	6	CA934243	CA934243 MTUITS.P5
C	40	8	5.4	196	6	CA934358	CA934358 MTUITS.P6
C	41	8	5.4	206	8	AZ788770	AZ788770 2M0036D12
C	42	8	5.4	226	9	CC795850	CC795850 SALK 0885
C	43	8	5.4	235	2	AW607315	AW607315 QV4-HT046
C	44	8	5.4	237	2	BB570424	BB570424 BB570424
C	45	8	5.4	237	8	AQ848625	AQ848625 LMAJFV1_1

RESULT 1

T38854/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

EST104418 S. cerevisiae strain X2180-1A

cDNA 3' end, mRNA sequence.

T38854

T38854.1

GI:622671

EST.

Saccharomyces cerevisiae (baker's yeast)

Saccharomyces cerevisiae

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

1 (bases 1 to 286)

Weinstock, K.

Saccharomyces cerevisiae cDNAs

Unpublished (1995)

Contact: Weinstock, K. and Venter, J.C.

The Institute for Genomic Research

932 Clopper Rd, Gaithersburg, MD 20878

Tel: 3018699056

Fax: 3018699423

Email: tdbinfo@db.tigr.org

For clone availability please contact the TIGR Database

(tdbinfo@db.tigr.org)

Seq primer: M13-21.

Location/Qualifiers

1. .286

/organism="Saccharomyces cerevisiae"

/mol_type="mRNA"

/strain="X2180-1A"

/db_xref="taxon:4932"

/clone_lib="S. cerevisiae strain X2180-1A"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:

FEATURES

source

ALIGNMENTS

T38854 286 bp mRNA linear EST 11-JAN-1995
EST104418 S. cerevisiae strain X2180-1A Saccharomyces cerevisiae
CDNA 3' end, mRNA sequence.

T38854
T38854.1 GI:622671

EST.
Saccharomyces cerevisiae (baker's yeast)

Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

1 (bases 1 to 286)

Weinstock, K.
Saccharomyces cerevisiae cDNAs

Unpublished (1995)
Contact: Weinstein, K. and Venter, J.C.

The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878

Tel: 3018699056
Fax: 3018699423

Email: tdbinfo@db.tigr.org
For clone availability please contact the TIGR Database
(tdbinfo@db.tigr.org)

Seq primer: M13-21.
Location/Qualifiers
1. .286

FEATURES

source
/organism="Saccharomyces cerevisiae"
/mol_type="mRNA"
/strain="X2180-1A"
/db_xref="taxon:4932"
/clone_lib="S. cerevisiae strain X2180-1A"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:

```

ORIGIN
Alignment Scores:
Pred. No.: 6.53 Length: 286
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.71% Indels: 0
DB: 7 Gaps: 0

US-09-851-138C-52 (1-149) x T38854 (1-286)

Qy 14 SerPheSerIlePheLeuLeuAlaLeuPhe 23
Db 160 TCCTTTTCGATATTCCTGCTGCATGTTT 131

RESULT 2
CL134102/c 1135 bp DNA linear GSS 05-JAN-2004
LOCUS ISB1-104H19_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-104H19,
DEFINITION genomic survey sequence.
ACCESSION CL134102
VERSION CL134102.1 GI:40627737
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE
1 (bases 1 to 1135)
Kremetzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: T7 TAATACGACTCACTATAGGG
Class: BAC ends
High quality sequence start: 95
High quality sequence stop: 215.
FEATURES
source
Location/Qualifiers
1..1135
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-104H19"
/clone_lib="ISB1"
/notes="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC
Library Segment 1"

ORIGIN
Alignment Scores:
Pred. No.: 21.4 Length: 1135
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.71% Indels: 0
DB: 9 Gaps: 0

US-09-851-138C-52 (1-149) x CL134102 (1-1135)

Qy 14 SerPheSerIlePheLeuLeuAlaLeuPhe 23
Db 583 TCCTTTTCGATATTCCTGCTGCATGTTT 554

RESULT 3
CC504025 181 bp DNA linear GSS 17-JUN-2003
LOCUS CH240_343J22.T7 CHORI-240 Bos taurus genomic clone CH240_343J22,
DEFINITION

```

```

genomic survey sequence.
CC504025 GI:31822318
VERSION CC504025.1
KEYWORDS GSS.
SOURCE Bos taurus (cow)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
REFERENCE
1 (bases 1 to 181)
Holt, R., Stott, J., Yang, G., Barber, S., Smailus, D., Prabhu, A.-L.,
Tsai, M., Cloutier, A., Lee, D., Girn, N., Olson, T., Mayo, M.,
Butterfield, Y., Kirkpatrick, R., Liu, J., Guin, R., Chan, A., Chiu, R.,
Mathewson, C., Wye, N., Masson, A., Brown-John, M., Jones, S.,
Schein, J., Marra, M., de Jong, P., McWilliam, S., Barris, W.,
Dalrymple, B. P. and Tellam, R.
Bovine BAC End Sequences from Library CHORI-240, PLATES 294 to 398
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: CH240_343J22.TARBAC13P2
Contact: Rob Holt
Sequencing
The British Columbia Cancer Agency Genome Science Centre
600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6
Tel: 604-877-6085
Fax: 604-877-6276
Email: rholt@bcgsc.ca
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering/information.htm). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the
British Columbia Genome Sciences Centre, Canada.
Plate: 343 row: J column: 22
Seq primer: T7
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..181
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_343J22"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/notes="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
Hereford bull LI Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

ORIGIN
Alignment Scores:
Pred. No.: 45.5 Length: 181
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.04% Indels: 0
DB: 9 Gaps: 0

US-09-851-138C-52 (1-149) x CC504025 (1-181)

Qy 92 ProCysAlaAlaThrAlaSerLeuArg 100
Db 88 CCCTGTGCGCTACAGCGCTCTGTGGA 114

RESULT 4
CD143793 286 bp mRNA linear EST 14-SEP-2003
LOCUS MGI-0088U-A295-H10-U.B MGI-0088 Schistosoma mansoni cDNA clone
DEFINITION MGI-0088U-A295-H10.B, mRNA sequence.
ACCESSION CD143793
VERSION CD143793.1 GI:34681617

```



```

KEYWORDS
SOURCE  EST.
ORGANISM Schistosoma mansoni

REFERENCE
AUTHORS Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,
Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,
Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
Sa.R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
TITLE Transcriptional analysis of the acoelomate human parasite Schistosoma
mansoni
JOURNAL Nat. Genet. 35 (2), 148-157 (2003)
MEDLINE 22879926
PUBMED 12973350
COMMENT Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjovski@usp.br
This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL http://bioinfo.iq.usp.br/schisto/
Plate: MGI-0088U-A295 row: 10 column: H.

FEATURES
source
1..286
Location/Qualifiers
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MGI-0088U-A295-H10.B"
/sex="mixed pool"
/dev_stage="germball"
/lab_host="Biomphalaria glabrata"
/clone_lib="MGI-0088"
/note="Vector: pGSM T-easy"

ORIGIN
Alignment Scores:
Pred. No.: 67.4 Length: 286
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.04% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-52 (1-149) x CD143793 (1-286)

Qy 19 LeuLeuAlaLeuPheSerCysLeuLeu 27
|||||
Db 95 TTACTAGTCCTCTCTCTCTCTCTCTCT 121

RESULT 5
AQ844124 312 bp DNA linear GSS 25-MAY-2001
LOCUS LMAJFV1_lm03e04.y2 Leishmania major FV1 random genomic library
DEFINITION Leishmania major genomic clone LMAJFV1_lm03e04 5', genomic survey
sequence.
ACCESSION AQ844124
VERSION AQ844124.1 GI:6011998
KEYWORDS GSS.
SOURCE Leishmania major
ORGANISM Leishmania major
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE
AUTHORS Akopyants,N.S., Clifton,S.W., Martin,J., Page,D., Wylie,T., Li,L.,
Kissinger,J.C., Roos,D.S. and Beverley,S.M.
TITLE A survey of the Leishmania major Friedlin strain V1 genome by
shotgun sequencing: a resource for DNA microarrays and expression
profiling
JOURNAL Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
MEDLINE 21192569
PUBMED 11295190
COMMENT Contact: Akopyants, NS / Beverley, SM
WashU Leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library construction: Natalia S. Akopyants, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
If using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
Friedlin strain V1 genome by shotgun sequencing' and the Washington
University Genome Sequencing Center For information on obtaining
clone material please contact: Natalia S. Akopyants Ph.D.
(natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
(beverley@borcim.wustl.edu)
Seq primer: -40RP from Gibco
Class: shotgun
High quality sequence stop: 278.
Location/Qualifiers
1..312
/organism="Leishmania major"
/mol_type="genomic DNA"
/strain="Friedlin strain V1"
/db_xref="taxon:5664"
/clone="LMAJFV1_lm03e04"
/lab_host="TOP10 (Invitrogen)"
/clone_lib="Leishmania major FV1 random genomic library"
/note="Vector: pZero-2 (Invitrogen); Site 1: EcoRV;
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
Phosphatase and ligated into pZero-2 vector's EcoRV site."

ORIGIN
Alignment Scores:
Pred. No.: 72.7 Length: 312
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.04% Indels: 0
DB: 8 Gaps: 0

US-09-851-138C-52 (1-149) x AQ844124 (1-312)

Qy 16 SerLePheLeuLeuAlaLeuPheSer 24
|||||
Db 179 TCCATCTTTTGTGGCGCTCTCTCC 205

RESULT 6
BH258325 349 bp DNA linear GSS 30-NOV-2001
LOCUS BH258325
DEFINITION CH230-60P6-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-60P6, genomic survey sequence.
ACCESSION BH258325
VERSION BH258325.1 GI:17160648
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 349)

```

AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P., and Fraser, C. M.
TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment
JOURNAL Unpublished (1999)
COMMENT Other GSSs: CH230-60P6.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the rat BAC library CHORI-230 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orering-information.htm>). BAC end page: http://www.tigr.org/tldb/bac_ends/rat/bac_end_intro.html
 Plate: 60 row: P column: 6
 Seq primer: T7
 Class: BAC ends.

FEATURES Location/Qualifiers
 1..349
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /strain="BN/SsNHsd/MCW"
 /db_xref="taxon:10116"
 /clone="CH230-60P6"
 /sex="Female"
 /cell_type="Brain"
 /clone_lib="CHORI-230 Segment 1"
 /notes="Vector: PTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI; CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by Pieter de Jong"

ORIGIN

Alignment Scores:
 Pred. No.: 80 Length: 349
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.04% Indels: 0
 DB: 8 Gaps: 0

US-09-851-138C-52 (1-149) x BH258325 (1-349)

QY 10 LeuProGlyCysSerPheSerlePhe 18
 |||||
 Db 244 CTACCAGGGTGCAGCTTCAGCAATTC 270

RESULT 7
 AQ845487
 LOCUS AQ845487
 DEFINITION LMAJFV1_lm26b07.y1 Leishmania major FV1 random genomic library
 Leishmania major genomic clone LMAJFV1_lm26b07 5', genomic survey sequence.
 ACCESSION AQ845487
 VERSION AQ845487.1 GI:6050135
 KEYWORDS GSS.
 SOURCE Leishmania major
 ORGANISM Leishmania major
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 1 (bases 1 to 372)
REFERENCE
AUTHORS Akopyants N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L., Kissinger, J.C., Roos, D.S. and Beverley, S.M.
TITLE A survey of the Leishmania major Friedlin strain V1 genome by shotgun sequencing: a resource for DNA microarrays and expression profiling
JOURNAL Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
MEDLINE 21192569
PUBMED 11295190

COMMENT Other GSSs: lm26b07.x1
 Contact: Akopyants, NS / Beverley, SM
 WashU Leishmania Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 Library construction: Natalia S. Akopyants, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 If using this information please cite:
 N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major Friedlin strain V1 genome by shotgun sequencing' and the Washington University Genome Sequencing Center for information on obtaining clone material please contact: Natalia S. Akopyants Ph.D. (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D. (beverley@borcim.wustl.edu)
 Seq primer: -40RP from Gibco
 Class: shotgun
 High quality sequence stop: 349.
 FEATURES Location/Qualifiers
 1..372
 /organism="Leishmania major"
 /mol_type="genomic DNA"
 /strain="Friedlin strain V1"
 /db_xref="taxon:5664"
 /clone="LMAJFV1_lm26b07"
 /lab_host="TOP10 (Invitrogen)"
 /clone_lib="Leishmania major FV1 random genomic library"
 /note="Vector: pZero-2 (Invitrogen); Site 1: EcoRV;
 Genomic DNA was isolated from stationary phase cells. For this library, DNA was sheared to give a tight size distribution of 1-1.5kb fragments, blunt-ended with T4 DNA polymerase, dephosphorylated with Shrimp Alkaline Phosphatase and ligated into pZero-2 vector's EcoRV site."

ORIGIN

Alignment Scores:
 Pred. No.: 84.5 Length: 372
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.04% Indels: 0
 DB: 8 Gaps: 0

US-09-851-138C-52 (1-149) x AQ845487 (1-372)

QY 16 SerlePheLeuLeuAlaLeuPheSer 24
 |||||
 Db 315 TCCATCTTTTGTGGCGCTCTCTCC 341

RESULT 8
 AI607477
 LOCUS AI607477
 DEFINITION mr84b07.y1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
 IMAGE:604117 5', mRNA sequence.
 ACCESSION AI607477
 VERSION AI607477.1 GI:4616644
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 413)
REFERENCE
AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Treising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)

MGI:369549

Seq primer: -40RP from Gibco

High quality sequence stop: 382

POLYA=No.

FEATURES source

Location/Qualifiers
 1. 413
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="NIH Swiss"
 /db_xref="taxon:10090"
 /clone="IMAGE:604117"
 /sex="pooled"
 /tissue_type="heart"
 /dev_stage="13 day embryos"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="Stratagene mouse heart (#937316)"
 /note="Organ: heart; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
 adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor
 sequence: 5' CTCGAGTTTTTTTTTTTTTT 3'"

ORIGIN

Alignment Scores:
 Pred. No.: 92.5 Length: 413
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.04% Indels: 0
 DB: 1 Gaps: 0

US-09-851-138C-52 (1-149) x AI607477 (1-413)

QY 16 SerilePheLeuAlaLeuPheSer 24
 DB 326 AGTATTTCTTACTTGCCTCTTTCT 352

RESULT 9

AI946531/c
 LOCUS bs27e02.y1 Drosophila melanogaster adult testis library Drosophila
 DEFINITION melanogaster cDNA clone bs27e02 5', mRNA sequence.
 ACCESSION AI946531
 VERSION AI946531.2 GI:9991854
 KEYWORDS EST.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 414)
 AUTHORS Andrew, J., Bouffard, G., Cheadle, C., Lu, J., Becker, K.G. and
 Oliver, B.
 TITLE Gene discovery using computational and microarray analysis of
 transcriptions in the drosophila melanogaster testis
 JOURNAL Genome Res. 10 (12), 2030-2043 (2000)
 MEDLINE 20568492
 PUBMED 11116097
 COMMENT On Aug 17, 1999 this sequence version replaced gi:5736957.
 Contact: Brian Oliver
 Laboratory of Cellular and Developmental Biology
 NIDDK, National Institutes of Health
 6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA

Fax: (301) 496 5239
 Email: oliver@helix.nih.gov,
 http://www.niddk.nih.gov/intram/people/boliver.htm
 Tissue isolation and library construction performed at the National
 Institute of Diabetes and Digestive and Kidney Diseases, NIH (see
 http://www.niddk.nih.gov/intram/people/boliver.htm). DNA sequencing
 and analyses performed by National Institutes of Health Intramural
 Sequencing Center (NISC; see http://www.nisc.nih.gov).
 Plate: 27 row: e column: 02
 Seq primer: M13RP1 reverse primer (ABI).

FEATURES

source

Location/Qualifiers

1. 414
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /strain="y[*] w[67cl]/y"
 /db_xref="taxon:7227"
 /clone="bs27e02"
 /sex="male"
 /dev_stage="1-5 day adult"
 /lab_host="SOLR (Stratagene)"
 /clone_lib="Drosophila melanogaster adult testis library"
 /note="Organ: testis; Vector: pBluescript SK (Stratagene);
 Site_1: EcoRI; Site_2: Xho I; Testes dissected from 1-5
 day adult y[*] w[67cl]/Y males raised at 25°C. RNA
 isolated using Trizol (Life Technologies) and a single
 round of Poly(A)+ selection using Oligotex (Qiagen). cDNA
 library constructed using Stratagene ZAP-cDNA synthesis
 kit. Oligo dt'-primed, size fractionated -1-6 kb, and
 directionally cloned at EcoRI and XhoI in Uni-ZAP XR.
 Following a single round of amplification pBluescript SK
 phagemids were mass excised. A distribution channel for
 clones is being sought, but not currently available.
 Requests for clones cannot be honored."

ORIGIN

Alignment Scores:
 Pred. No.: 92.7 Length: 414
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.04% Indels: 0
 DB: 1 Gaps: 0

US-09-851-138C-52 (1-149) x AI946531 (1-414)

QY 25 CysLeuLeuThrProThrAlaGlyLeu 33

DB 283 TGCCTACTCTCCACAGCTGGTCTT 257

RESULT 10

CO299961/c
 LOCUS EX177107.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
 DEFINITION melanogaster cDNA clone EX177107 5, mRNA sequence.
 ACCESSION CO299961
 VERSION CO299961.1 GI:49221657
 KEYWORDS EST.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 446)
 AUTHORS Kopczyński, C., Platt, D., Campbell, J., Muzong, C., Laufer, A.,
 Peterson, E. and Swimmer, C.
 TITLE Exelixis FlyTag EST Project CK01 Library
 JOURNAL Unpublished (2004)
 COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

```

Plate: EK.1771 row: A column: 7
High quality sequence stop: 440.
Location/Qualifiers
1. 446
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clones="EK177107"
/clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
/note="Organ: mixed stage embryos, imaginal disks, and
adult heads; Vector: pCDNA-SK+; Site: 1: NotI; Site 2:
XhoI; Random primed, normalized library from mixed stage
embryos, imaginal disks, and adult heads."

ORIGIN
Alignment Scores:
Pred. No.: 98.8 Length: 446
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservat: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.04% Indels: 0
DB: 7 Gaps: 0

US-09-851-138C-52 (1-149) x CO299961 (1-446)

Qy 15 PheSerIlePheLeuLeuAlaLeuPhe 23
|||||
Db 350 TTCTCCATTTTCTGCTTGCTCTATT 324
|||||

RESULT 11
LOCUS CO325948 456 bp mRNA linear EST 28-JUN-2004
DEFINITION EP02947.3prime Exelixis FlyTag CK02 pCDNA-SK+ Drosophila
melanogaster cDNA clone EP02947 3, mRNA sequence.
ACCESSION CO325948
VERSION CO325948.1 GI:49384382
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 456)
AUTHORS Platt,D., Kopczyński,C., Muzong,C., Laufer,A., Leung,W.,
Peterson,E. and Swimmer,C.
TITLE Exelixis FlyTag EST Project CK02 Library
JOURNAL Unpublished (2004)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: Ep.29 row: D column: 11
High quality sequence stop: 455.
Location/Qualifiers
1. 456
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="EP02947"
/clone_lib="Exelixis FlyTag CK02 pCDNA-SK+"
/note="Organ: mixed stage embryos, imaginal disks, and
adult heads; Vector: pCDNA-SK+; Site: 1: NotI; Site 2:
XhoI; Random primed, normalized library from mixed stage
embryos, imaginal disks, and adult heads. Subset of
Exelixis FlyTag CK01 clones sequenced from 3' end"

ORIGIN
Alignment Scores:
Pred. No.: 101 Length: 456
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservat: 0

```

```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.04% Indels: 0
DB: 7 Gaps: 0

US-09-851-138C-52 (1-149) x CO325948 (1-456)

Qy 15 PheSerIlePheLeuLeuAlaLeuPhe 23
|||||
Db 289 TTCTCCATTTTCTGCTTGCTCTATT 315
|||||

RESULT 12
LOCUS H94357 477 bp mRNA linear EST 25-NOV-1996
DEFINITION YW54C05.r1 Soares_placenta_8to9weeks_2NBHP8to9W Homo sapiens cDNA
clone IMAGE:256040 5' similar to gb:MI4565 CYTOCHROME P450 XI1A1,
MITOCHONDRIAL (HUMAN);, mRNA sequence.
ACCESSION H94357
VERSION H94357.1 GI:1101990
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 477)
AUTHORS Hillier,L., Clark,N., Dubucque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Treviski,S., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 371
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 788 Std Error: 0.00
Seq primer: M13RP1.
Location/Qualifiers
1. 477
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3885650"
/db_xref="taxon:9606"
/clone="IMAGE:256040"
/dev_stages="two placentae: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DHI0B (ampicillin resistant)"
/clone_lib="Soares_placenta_8to9weeks_2NBHP8to9W"
/note="Organ: placenta; Vector: p7T73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCGATTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."

ORIGIN
Alignment Scores:
Pred. No.: 105 Length: 477
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservat: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.04% Indels: 0
DB: 7 Gaps: 0

```

```

US-09-851-138C-52 (1-149) x H94357 (1-477)
Qy 61 ILeuHisLeuProGlyCysValpro 69
Db 425 ATCTGCACCTTCCAGGTTGGTGCCA 399

RESULT 13
BF442449 484 bp mRNA linear EST 01-DEC-2000
DEFINITION 259170 MARC 2Pig Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF442449
VERSION BF442449.1 GI:11502541
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 484)
Vallet, J., Wise, T., Rohrer, G.A., Perte, G., Sultana, R.,
Fahrenkrug, S.C., Smith, T.P.L., Freking, B.A., Cho, J., White, J.,
Quackenbush, J. and Keefe, J.W.
Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
Mamm. Genome 13 (8), 475-478 (2002)
22213789
12226715
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGCG
Plate: 89 row: N column: 20
Seq primer: ATTGAGTGACACTATAG.
FEATURES
source
1..484
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 2Pig"
/note="Vector: pCMV SPORR6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
ORIGIN
Alignment Scores:
Pred. No.: 106 Length: 484
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.04% Indels: 0
DB: 2 Gaps: 0
US-09-851-138C-52 (1-149) x BF442449 (1-484)
Qy 26 LeuLeuThrProThrAlaGlyLeuGlu 34
Db 424 CTTCTGACCACTGCGAGTTGGAG 450

RESULT 14
CL369856 522 bp DNA linear GSS 19-AUG-2004
LOCUS RPCI44_292B18.f RPCI-44 Sus scrofa genomic clone RPCI44_292B18,
genomic survey sequence.
ACCESSION CL369856
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;

```

```

VERSION CL369856.1 GI:51421821
KEYWORDS GSS.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 522)
Rogatcheva, M.B., Meyers, S., He, W., Larkin, D.M., Marron, B.M.,
Beever, J.E. and Schook, L.B.
Piggy-BACing the Human Genome: Constructing a Porcine Physical Map
Through Comparative Genomics
Unpublished (2004)
Other GSSs: RPCI44_292B18.f
Contact: Lawrence B. Schook
Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 265 5326
Fax: 217 244 5617
Email: schook@uiuc.edu
Clones are derived from the porcine BAC library RPCI-44
(http://www.bacpac.chori.org/porcine242.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@chori.org).
Clones may be purchased from BACPAC Resources
(http://BACPACresources.chori.org). This work was undertaken as part
of the International Swine Genome Sequencing Consortium by
University of Illinois at Urbana Champaign, USA with funds provided
by grant No. AG2002-34480-11828 from USDA-CSREES and
AG2001-35205-09965 from USDA/NRI (Livestock Genome Sequencing
Initiative)
Plate: 292 row: B column: 18
Seq primer: T7
Class: BAC ends.
FEATURES
source
1..522
/organism="Sus scrofa"
/mol_type="genomic DNA"
/strain="four pigs (breed: 37.5% Yorks Landrace and 25%
Meishan)"
/db_xref="taxon:9823"
/clone="RPCI44_292B18"
/sex="male"
/cell_type="blood"
/clone_lib="RPCI-44"
/note="Vector: pTARBAC2; Site 1: EcoRI; Site 2: EcoRI;
porcine male BAC library produced by Pieter de Jong"
ORIGIN
Alignment Scores:
Pred. No.: 113 Length: 522
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.04% Indels: 0
DB: 9 Gaps: 0
US-09-851-138C-52 (1-149) x CL369856 (1-522)
Qy 26 LeuLeuThrProThrAlaGlyLeuGlu 34
Db 371 CTTCTGACCACTGCGAGTTGGAG 397

RESULT 15
BQ039335 549 bp mRNA linear EST 17-APR-2002
LOCUS Gd09a08.y1 Moss EST library PPS Physcomitrella patens cDNA clone
DEFINITION PEP SOURCE_ID:PPS30116 5', mRNA sequence.
ACCESSION BQ039335
VERSION BQ039335.1 GI:19778637
KEYWORDS EST.
SOURCE Physcomitrella patens
ORGANISM Physcomitrella patens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;

```

Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.

1 (bases 1 to 549)

Quatrano,R., Bashardes,S., Cove,D., Cumming,A., Knight,C., Clifton,S., Marra,L., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Steptoe,M., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.

Leeds/Wash U Moss EST Project

Unpublished (1999)

Other ESTs: gd09a08.x1

Contact: Ralph Quatrano

Leeds/Wash U Moss EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Libraries were constructed by Dr. Stavros Bashardes as part of the Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and Washington Univ. in St. Louis (USA) DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)

Seq primer: -40RP from Gibco

High quality sequence stop: 422.

Location/Qualifiers

1..549

/organism="Physcomitrella patens"

/mol_type="mRNA"

/db_xref="taxon:3218"

/clone="PEP_SOURCE.ID:PPS30116"

/dev_stage="protonemata, 7day old untreated"

/lab_host="E.coli DH10b"

/clone_lib="Moss EST library PPS"

/notes="Vector: pBluescript SK-; Site 1: XhoI; Site 2: EcoRI; Library constructed by Stavros Bashardes and re-arrayed by A. Cumming & Honglin Rong. Construction of the cDNA library was carried out using Statagene's 'Unizap' - cDNA synthesis kit, to ligate cDNA directionally in Unizap XR vector arms. The vector is designed containing the pBluescript sequence as well as the lambda DNA and cDNA is cloned in the EcoRI and XhoI sites in the pBluescript sequence. The vector was then packaged using Gold gigapackaging extracts, propagated in XL-IBLue MRF cells and amplified. The library was excised by mass excision using Statagene's Mass excision kit to infect SOUK cells with phagemids and ampicillin resistant transformants selected. Approximately 1,000,000 colonies were grown and recovered by using Quiagen midi prep kit.2 micro grams of plasmid DNA were used to transform DH10b cells by electroporation. Clones corresponding to abundant transcripts were identified by colony hybridization and eliminated from the library, be rearraying. This library is non-directionally cloned."

ORIGIN

Alignment Scores:

Pred. No.:	118	Length:	549
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	6.04%	Indels:	0
DB:	5	Gaps:	0

US-09-851-138C-52 (1-149) x BQ039335 (1-549)

QY 71 ValArgSerGlyAsnThrSerArgCys 79
 DB 428 GTCCGATCGGAAACACGAGTCGTTGT 454

Search completed: March 3, 2005, 21:58:07
 Job time : 9394.29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:41:12 ; Search time 469.169 Seconds
(without alignments)
1239.345 Million cell updates/sec

Title: US-09-851-138C-138
Perfect score: 12
Sequence: 1 LEYRNAGLYMV 12

Scoring table:

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9403671

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEVS=xlp
-O=/cgn2_1/USPTO_spool_p/US09851138/runat_28022005_120306_21465/app_query.fasta_1.1123
-DB=GenEmbl -OPMT=fastap -SUFFIX=olig.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09851138 @CGN 1 1 6331 @runat_28022005_120306_21465 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	100.0	447	6 A50396	A50396 Sequence 51
2	12	100.0	447	6 AR127536	AR127536 Sequence
3	12	100.0	447	14 HPC00REAL	L39317 Hepatitis C
4	12	100.0	1584	14 HPCJK030A3	D49747 Hepatitis C

5	11	91.7	867	14	HPCHC048
6	10	83.3	411	14	HPCCP3
7	10	83.3	1584	14	HPCJK049A5
8	10	83.3	1587	14	HPC00RE02
9	10	83.3	1794	14	AY231584
10	10	83.3	1794	14	AY231585
11	10	83.3	1794	14	AY231587
12	10	83.3	1799	14	AY231586
13	10	83.3	1799	14	AY231588
14	10	83.3	1799	14	AY231589
15	10	83.3	1799	14	AY231590
16	10	83.3	9444	6	E10839
17	10	83.3	9444	6	E10841
18	10	83.3	9444	14	HPCFG
19	10	83.3	9450	14	HPCJK049E1
20	8	66.7	1504	14	HPCNE145G
21	8	66.7	1584	14	HPCJK055A6
22	8	66.7	9444	6	E10840
23	8	66.7	143957	9	AL356313
24	8	66.7	243049	2	AC095808
25	8	66.7	301726	2	AC133666
26	8	66.7	311323	2	AC098437
27	7	58.3	231	12	AY658374
28	7	58.3	352	14	AY177857
29	7	58.3	352	14	AY177874
30	7	58.3	438	11	G84333
31	7	58.3	474	14	AY739371
32	7	58.3	474	14	AY739396
33	7	58.3	474	14	AY739398
34	7	58.3	474	14	AY739400
35	7	58.3	474	14	AY739412
36	7	58.3	566	11	G79549
37	7	58.3	566	11	G93318
38	7	58.3	576	14	AF134737
39	7	58.3	577	6	E04081
40	7	58.3	699	6	AR507465
41	7	58.3	1280	14	AY746873
42	7	58.3	1280	14	AY746874
43	7	58.3	1280	14	AY746875
44	7	58.3	1280	14	AY746876
45	7	58.3	1280	14	AY746877

ALIGNMENTS

RESULT 1	A50396	Sequence 51 from Patent WO9613590.	447 bp	DNA	linear	PAT 07-MAR-1997
LOCUS	A50396					
DEFINITION	A50396					
ACCESSION	A50396					
VERSION	A50396.1	GI:2303407				
KEYWORDS						
SOURCE	unidentified					
ORGANISM	unidentified					

REFERENCE 1 (bases 1 to 447)

Maertens,G. and Stuyver,L.

NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS

JOURNAL Patent: WO 9613590-A 51 09-MAY-1996

INNOGENETICS NV (BE)

Other publication AU 3844095 960523.

Location/Qualifiers

source 1..447

/organism="unidentified"

/mol_type="unassigned DNA"

/db_xref="taxon:32644"

ORIGIN

Alignment Scores:	0.000199	Length:	447
Pred. No.:	12.00	Matches:	12
Score:	100.00%	Conservative:	0

```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-138 (1-12) x A50396 (1-447)

QY 1 LeuGluTyrArgAsnAlaSerGlyLeuTyrMetVal 12
   |||||
Db 97 CTGGAGTACCGTAATGCTCCGGACTCTACATGTA 132

RESULT 2
LOCUS AR127536 447 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 51 from patent US 6180768.
ACCESSION AR127536
VERSION AR127536.1 GI:14114131
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS Maertens,G. and Stuyver,L.
TITLE Sequences of hepatitis C virus genotypes and their use as
        prophylactic, therapeutic and diagnostic agents
JOURNAL Patent: US 6180768-A 51 30-JAN-2001;
FEATURES
        Location/Qualifiers
        1..447
        /organism="unknown"
        /mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.: 0.000199 Length: 447
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-138 (1-12) x AR127536 (1-447)

QY 1 LeuGluTyrArgAsnAlaSerGlyLeuTyrMetVal 12
   |||||
Db 97 CTGGAGTACCGTAATGCTCCGGACTCTACATGTA 132

RESULT 3
HPCOREAL
LOCUS Hepatitis C virus type 3 clone NL96 precursor protein gene, partial
DEFINITION
ACCESSION L39317 GI:845497
VERSION L39317.1
KEYWORDS
SOURCE Hepatitis C virus type 3
ORGANISM Hepatitis C virus
REFERENCE
AUTHORS van Doorn,L.J., Kleter,B., Stuyver,L., Maertens,G., Brouwer,H.,
        Schalm,S., Heijink,R. and Quint,W.
TITLE Analysis of hepatitis C virus genotypes by a line probe assay and
        correlation with antibody profiles
JOURNAL J. Hepatol. 21 (1), 122-129 (1994)
MEDLINE 95052487
PUBMED 7525693
REFERENCE
AUTHORS van Doorn,L.J., Kleter,G.E., Stuyver,L., Maertens,G., Brouwer,J.T.,
        Schalm,S.W., Heijink,R.A. and Quint,W.G.
TITLE Sequence analysis of hepatitis C virus genotypes 1 to 5 reveals
        multiple novel subtypes in the Benelux countries
JOURNAL J. Gen. Virol. 76 (Pt 7), 1871-1876 (1995)
MEDLINE 97201609
PUBMED 9049395

FEATURES
        source
        Location/Qualifiers
        1..447
        /organism="Hepatitis C virus"
        /mol_type="genomic RNA"
        /db_xref="taxon:40363"
        /clone="NL96"
        /notes="genotype: 3"
        <1..>447
        /codon_start=1
        /product="precursor protein"
        /protein_id="AA67840.1"
        /db_xref="GI:845498"
        /translation="DCINPATNLPGCSFIPLALFSCLLTPTAGLEYNASGLYMW
        TNDCSGSIYVAGDIILHLPGCVPSRGNTSRWIPVSPVAVKSPCAATASLRTH
        VDMVGAATLCSALYVGDLCGALFLVGQGSFWRHQHWTVQDCNCI"
        1..96
        /product="core protein"
        /note="putative"
        97..447
        /product="e1 protein"
        /note="putative"

mat_peptide
mat_peptide

ORIGIN
Alignment Scores:
Pred. No.: 0.000199 Length: 447
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-851-138C-138 (1-12) x HPCOREAL (1-447)

QY 1 LeuGluTyrArgAsnAlaSerGlyLeuTyrMetVal 12
   |||||
Db 97 CTGGAGTACCGTAATGCTCCGGACTCTACATGTA 132

RESULT 4
HPCJK030A3
LOCUS Hepatitis C virus isolate JK030 gene for core, env, and part of
DEFINITION E2/NS1, partial cds.
ACCESSION D49747 GI:1197102
VERSION D49747.1
KEYWORDS core, env, and part of E2/NS1.
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
REFERENCE
AUTHORS Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tenda,F.,
        Lesmana,L.A., Miyakawa,Y. and Mayumi,M.
TITLE Hepatitis C virus variants from Jakarta, Indonesia classifiable
        into novel genotypes in the second (2e and 2f), tenth (10a) and
        eleven (11a) genetic groups
JOURNAL J. Gen. Virol. 77 (Pt 2), 293-301 (1996)
MEDLINE 96226020
PUBMED 8627233
REFERENCE
AUTHORS Okamoto,H.
TITLE Unpublished
JOURNAL
AUTHORS Okamoto,H.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1995) Hiroaki Okamoto, Jichi Medical School,
        Immunology Division, Minamikawachi-machi, Kawachi-gun, Tochigi
        329-04, Japan (E-mail:hokamoto@jichi.ac.jp,
        Tel:0285-44-2111(ex.3334), Fax:0285-44-1557)
        Location/Qualifiers
        1..1584
        /organism="Hepatitis C virus"
        /mol_type="genomic RNA"
        /isolate="JK030"

```



```
/db_xref="taxon:11103"
<1..339
/citation=[2]
/evidence=not_experimental
340..>1584
/citation=[2]
/codon_start=1
/evidence=not_experimental
/product="core, env, and part of E2/NS1"
/protein_id="BAA08581.1"
/db_xref="GI:1197103"
/translation="MSLTPKQRIKNTNRPDVKFPGGQIVGGVYVLPREGPKL
GVRAVKTSRQSRSRQPIPRARTREGSWAQGFPMPLIGNEGGNAGWLLSPRG
SRPAGNDPRRRNRNLGKVIDTLTCGFADLMGYIPLVGAPEGVARALAHIRALE
GINFENLPCFCSIFLLALLSLTPAAGLEYRNASGLYMTVNDGNSIVYEAGD
IILPLGCVPCVRSGNTSKWTSVPTVAVSHPGAATASLTHVDMVMVGAATLCSALY
VGDLGALFLVGGGFSKRRROHTVQDCNCSIYPGHLTGHRMAMNMNNSPATLVV
SQVRLPQTLLDVLGAHWGMAGVAYYSMQGNWAKVPLVLCLFSGVDASTRISGGA
AHNTYGLSLFSSGPKQNIQIN"

ORIGIN
Alignment Scores:
Pred. No.: 0.0056 Length: 1584
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-851-138c-138 (1-12) x HPCJK030A3 (1-1584)

Qy 1 LeuGlutYrArGAsnAlaSerGlyLeuTyMetVal 12
Db 913 TTGGAGTACCGTAATGCTCTCGGACTCTACATGGTA 948

RESULT 5
HPCHCV048 HPCHCV048 867 bp RNA linear VRL 15-FEB-2003
LOCUS Hepatitis C virus DNA, clone:BA-1.
DEFINITION Hepatitis C virus
ACCESSION D16736
VERSION D16736.1 GI:506254
KEYWORDS Hepatitis C virus
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.

REFERENCE
1 Ohno,T.
AUTHORS Ohno,T.
TITLE Hepatitis C virus
JOURNAL Thesis (1993) The University of Tokyo
REFERENCE 2 (bases 1 to 867)
AUTHORS Ohno,T.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-1993) Tomoyoshi Ohno, Nagoya City University
Medical School, Second Department of Internal Medicine; 1-1
Kavasaki, Mizuho, Nagoya, Aichi 467, Japan
(Tel:052-851-5511(ex.8748,2265), Fax:052-852-0849)

FEATURES
source
Location/Qualifiers
1..867
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/db_xref="taxon:11103"
/clone="BA-1"

ORIGIN
Alignment Scores:
Pred. No.: 0.00436 Length: 867
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.67% Indels: 0
DB: 14 Gaps: 0

US-09-851-138c-138 (1-12) x HPCHCV048 (1-867)

Qy 1 LeuGlutYrArGAsnAlaSerGlyLeuTyMet 11
Db 461 CTGGAGTACAGAACGCGTCGGGCTATACATG 493

RESULT 6
HPCCP3 HPCCP3 411 bp RNA linear VRL 07-FEB-1999
LOCUS Hepatitis C virus (individual isolate Td-3/93) gene for polyprotein
DEFINITION Hepatitis C virus (core protein (carboxy terminus) and E1
precursor, partial cds
envelope protein (amino terminus half)).
ACCESSION D30046
VERSION D30046.1 GI:485798
KEYWORDS E1 envelope protein; core protein.
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.

REFERENCE
1 (sites)
AUTHORS Hotta,H., Handajani,R., Lusida,M.I., Soemarto,W., Doi,H.,
Miyajima,H. and Homma,M.
SUBTYPE analysis of hepatitis C virus in Indonesia on the basis of
NS5b region sequences
J. Clin. Microbiol. 32 (12), 3049-3051 (1994)
95189942
PUBMED 7883898
REFERENCE 2 (bases 1 to 411)
AUTHORS Hotta,H.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 411)
AUTHORS Hotta,H.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1994) Hak Hotta, Kobe University School of
Medicine, Department of Microbiology; 7-5-1 Kusunoki-cho, Chuo-ku,
Kobe, Hyogo 650, Japan (Tel:078-341-7451(ex.3301),
Fax:078-351-6347)
Submitted (28-Apr-1994) to DDBJ by:
Hak Hotta
Kobe University School of Medicine
Department of Microbiology
7-5-1 Kusunoki-cho, Chuo-ku
Kobe, Hyogo 650
Japan
Phone: 078-341-7451 x3301
Fax: 078-351-6347.

FEATURES
source
Location/Qualifiers
1..411
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="Td-3/93"
/db_xref="taxon:11103"
<1..>411
/note="The carboxy terminus of the core protein and the
amino terminal half of the E1 envelope protein of
hepatitis C virus"
/codon_start=1
/product="polyprotein precursor"
/protein_id="BAA06282.1"
/db_xref="GI:485799"
/translation="LSCLLTPTAGLEYRNASGLYIVTNDGNSIVVEAQDIILHMPG
CDPCVRSNTSRCTVPSVSTVAVRAGAATASLTHVDMVMVGAATLCSALYVGLCGA
LFLVGGGFSWRHRQHTVQDCNCSIYPGHLTGHRM"

ORIGIN
Alignment Scores:
Pred. No.: 0.0302 Length: 411
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 83.33% Indels: 0
DB: 14 Gaps: 0
```

US-09-851-138C-138 (1-12) x HPCCF3 (1-411)

Qy 1 LeuGlutTyzArgAsnAlaSerGlyLeuTy 10
 Db 31 CTGGAGTATCGTAATGCTCCGGACTCTAC 60

RESULT 7

HPCJK049A5

LOCUS

DEFINITION Hepatitis C virus isolate JK049 RNA linear VRL 10-FEB-1999
 E2/NS1, partial cds.

ACCESSION

D49749.1 GI:1197110

VERSION

D49749.1 GI:1197110

KEYWORDS

core, env, and part of E2/NS1.

SOURCE

Hepatitis C virus

ORGANISM

Hepatitis C virus

REFERENCE 1 (bases 1 to 1584)
 Tokita,H., Okamoto,H., Izuka,H., Kishimoto,J., Tsuda,F.,
 Lesmana,L.A., Miyakawa,Y. and Mayumi,M.
 Hepatitis C virus variants from Jakarta, Indonesia classifiable
 into novel genotypes in the second (2e and 2f), tenth (10a) and
 eleventh (11a) genetic groups

J. Gen. Virol. 77 (Pt 2), 293-301 (1996)

JOURNAL

MEDLINE

6226020

PUBMED

8627233

REFERENCE 2 (bases 1 to 1584)

Okamoto,H.

AUTHORS

Unpublished

JOURNAL

REFERENCE 3 (bases 1 to 1584)

Okamoto,H.

AUTHORS

Direct Submission

TITLE

Submitted (17-MAR-1995) Hiroaki Okamoto, Jichi Medical School,
 Immunology Division; Minamikawachi-machi, Kawachi-gun, Tochigi
 329-04, Japan (E-mail:hokamoto@jichi.ac.jp,
 Tel:0285-44-2111(ex.3334), Fax:0285-44-1557)

JOURNAL

FEATURES

source

1. 1584

/organism="Hepatitis C virus"

/mol_type="genomic RNA"

/isolate="JK049"

/db_xref="taxon:11103"

<1..339

/citation=[2]

/evidence=not_experimental

340..>1584

/note="core, env, and part of E2/NS1"

/protein_id="BAA08593.1"

/db_xref="GI:1197111"

/translations="MSTLPKQRTKRNTPRRPQNVKFPGGQIVGGVTVLPRGPKL
 GVRVKTSSQPRRRQPIKARTEGSRWAQPGYPLVGNCGWAGWLLSPRG
 SRPSGNDPRRRNLGKVIDTLTCGFADLMGYIPLVGPVGVARALAHGVRALED
 GINFTATGNLPGCSFIFLLALLSCLLTPTAGLEYNASGLYTVTNDSCNSGIVYEAGD
 VIHLPGCTPCVLNNASKCTVPTVAVSRFGAATSLRTHVDMVGAATLCSALY
 VGLDGLGALFLVGGQFSWRHQHWTVQDCNCSLYPGHLTGHRMADMMNWSPATLIV
 SOVLRLPQTMFDLVI GAHWGVAGVAVYVSMQGNWAKVFLVLCIFSGVDASTITGGVA
 ASGAFITISLFSGAKQPLHLVN"

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

Length:

Matches:

Conservative:

Mismatch:

Indels:

Gaps:

US-09-851-138C-138 (1-12) x HPCJK049A5 (1-1584)

Qy 1 LeuGlutTyzArgAsnAlaSerGlyLeuTy 10
 Db 913 CTGGAGTATCGTAATGCTCCGGACTCTAC 942

RESULT 8

HPCCORE02

LOCUS

DEFINITION Hepatitis C virus gene for core, env, and part of E2/NS1, partial
 cds.

ACCESSION

D37839

VERSION

D37839.1 GI:1009250

KEYWORDS

Hepatitis C virus

SOURCE

Hepatitis C virus

ORGANISM

Hepatitis C virus

REFERENCE 1 (bases 1 to 1587)
 Tokita,H., Okamoto,H., Luengrojanakul,P., Vareesangthip,K.,
 Chainuvati,T., Izuka,H., Tsuda,F., Miyakawa,Y. and Mayumi,M.
 Hepatitis C virus variants from Thailand classifiable into five
 novel genotypes in the sixth (6b), seventh (7c, 7d) and ninth (9b,
 9c) major genetic groups

J. Gen. Virol. 76 (Pt 9), 2329-2335 (1995)

JOURNAL

MEDLINE

96005057

PUBMED

7561773

REFERENCE 2 (bases 1 to 1587)

Okamoto,H.

AUTHORS

Direct Submission

TITLE

Submitted (02-AUG-1994) Hiroaki Okamoto, Jichi Medical School,
 Immunology Division; Minamikawachi-machi, Kawachi-gun, Tochigi
 329-04, Japan (E-mail:hokamoto@jichi.ac.jp,
 Tel:0285-44-2111(ex.3334), Fax:0285-44-1557)

JOURNAL

FEATURES

Location/Qualifiers

1. 1587

/organism="Hepatitis C virus"

/mol_type="genomic RNA"

/isolate="Th527"

/db_xref="taxon:11103"

<1..339

/citation=[1]

/evidence=not_experimental

340..>1587

/note="core, env, and part of E2/NS1"

/citation=[1]

/codon_start=1

/evidence=not_experimental

/product="polyprotein"

/protein_id="BAA07086.1"

/db_xref="GI:1009251"

/translation="MSTLPKQRTKRNTPRRPQNVKFPGGQIVGGVTVLPRGPKL
 GVRVKTSSQPRRRQPIKARTEGSRWAQPGYPLVGNCGWAGWLLSPRG
 SRPSGNDPRRRNLGKVIDTLTCGFADLMGYIPLVGPVGVARALAHGVRALED
 GINFTATGNLPGCSFIFLLALLSCLLTPTAGLEYNASGLYTVTNDSCNSGIVYEAAE
 VIHLPGCTPCVLNNASKCTVPTVAVSRFGAATSLRTHVDMVGAATLCSALY
 VGLDGLGALFLVGGQFSWRHQHWTVQDCNCSLYPGHVSGRHMDMMNWSPAVLV
 SHLLRLPQTMFDLVI GAHWGVAGVAVYVSMQGNWAKVCIIVLMFSGVDKGTHTTGGT
 MASGANILNKLFPSCQQNLQVNV"

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

Length:

Matches:

Conservative:

Mismatch:

Indels:

Gaps:

US-09-851-138C-138 (1-12) x HPCCORE02 (1-1587)

Qy 1 LeuGlutTyzArgAsnAlaSerGlyLeuTy 10

Db 913 CTAGAGTACAGGACGCGCTCCGGCTTATAC 942

```

RESULT 9
AY231584 LOCUS AY231584 1794 bp RNA linear VRL 01-MAR-2004
DEFINITION Hepatitis C virus isolate NB179 polyprotein gene, partial cds.
ACCESSION AY231584
VERSION AY231584.1 GI:37961929
KEYWORDS
SOURCE
ORGANISM
Hepatitis C virus
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE
1 (bases 1 to 1794)
Chaudhuri,S. and Naik,T.N.
Molecular epidemiology of HCV infection among acute and chronic
liver disease patients in Kolkata, India
Unpublished
JOURNAL
AUTHORS
2 (bases 1 to 1794)
Chaudhuri,S. and Naik,T.N.
Direct Submission
JOURNAL
TITLE
Submitted (07-FEB-2003) Division of Virology, National Institute of
Cholera and Enteric Diseases, P-33 CIT Road, Scheme- XM,
Beliaghata, Kolkata, WB 700010, India
FEATURES
Location/Qualifiers
source
1..1794
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="NB179"
/db_xref="taxon:11103"
/note="genotype: 3b"
254..>1794
/note="contains core and envelope proteins"
/contig_start=1
/product="polyprotein"
/protein_id="AAP69955.1"
/db_xref="GI:37961930"
/translacion="MSTLPKPKROTENTLRPKKVPKAGQIVGVVLPKPPQPL
GVREVCAYEWSQPRGRTPTKAPREGSWAQGYWPVLYGEGCGWAGWLLPPRG
SRPWSQNDPRRRNRNGKVIDTLCGFADLMGYIPLIGAPVGGVARALAHGVRLED
GVNATGNLPCGFSIFLLALFSLCTCPASSLEYRNAGSLYLLTNDGNSRISVYEAD
VILHPCGVCVETDNNNTSCWTPISPTVAVKHGVTAGIRNHNMLVAPPTLCSAL
YVEDAFGASVILGQAFTRPROHKTVOCTNCISIVPGHVSGRHMAWMMWSPALGLV
ISHLNRLPQTFDLVGAHWGMAGLAYFSMQGNWAKVIVILMFSGVDATHTHTGGS
AAQTAGTSTFTFRGSPQNLQVNSNGSWHNSLTALNCNDSLTGTAGLFIYHKFNS
SGCPRMSCKPITYFNQGWGLPTDANINGPSDRPYCMHYPPRPNCITKPLNARYI
ASLPAP"
ORIGIN
Alignment Scores: 0.101 Length: 1794
Pred. No.: 10.00 Matches: 10
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 83.33% Gaps: 0
DB: 14
US-09-851-138C-138 (1-12) x AY231585 (1-1794)
QY 1 LeuGluTyArgAsnAlaSerGlyLeuTyr 10
Db 827 CTGGAGTACAGGATGCTGCGCTATAT 856
RESULT 11
AY231587 LOCUS AY231587 1794 bp RNA linear VRL 01-MAR-2004
DEFINITION Hepatitis C virus isolate 42 polyprotein gene, partial cds.
ACCESSION AY231587
VERSION AY231587.1 GI:37961935
KEYWORDS
SOURCE
ORGANISM
Hepatitis C virus
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE
1 (bases 1 to 1794)
Chaudhuri,S. and Naik,T.N.
Molecular epidemiology of HCV infection among acute and chronic
liver disease patients in Kolkata, India
Unpublished
JOURNAL
AUTHORS
2 (bases 1 to 1794)
Chaudhuri,S. and Naik,T.N.
Direct Submission
JOURNAL
TITLE
Submitted (07-FEB-2003) Division of Virology, National Institute of
Cholera and Enteric Diseases, P-33 CIT Road, Scheme- XM,
Beliaghata, Kolkata, WB 700010, India
FEATURES
Location/Qualifiers
source
1..1794
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="NB179"
/db_xref="taxon:11103"
/note="genotype: 3b"
254..>1794
/note="contains core and envelope proteins"
/contig_start=1
/product="polyprotein"
/protein_id="AAP69955.1"
/db_xref="GI:37961930"
/translacion="MSTLPKPKROTENTLRPKKVPKAGQIVGVVLPKPPQPL
GVREVCAYEWSQPRGRTPTKAPREGSWAQGYWPVLYGEGCGWAGWLLPPRG
SRPWSQNDPRRRNRNGKVIDTLCGFADLMGYIPLIGAPVGGVARALAHGVRLED
GVNATGNLPCGFSIFLLALFSLCTCPASSLEYRNAGSLYLLTNDGNSRISVYEAD
VILHPCGVCVETDNNNTSCWTPISPTVAVKHGVTAGIRNHNMLVAPPTLCSAL
YVEDAFGASVILGQAFTRPROHKTVOCTNCISIVPGHVSGRHMAWMMWSPALGLV
ISHLNRLPQTFDLVGAHWGMAGLAYFSMQGNWAKVIVILMFSGVDATHTHTGGS
AAQTAGTSTFTFRGSPQNLQVNSNGSWHNSLTALNCNDSLTGTAGLFIYHKFNS
SGCPRMSCKPITYFNQGWGLPTDANINGPSDRPYCMHYPPRPNCITKPLNARYI
ASLPAP"
ORIGIN
Alignment Scores: 0.101 Length: 1794
Pred. No.: 10.00 Matches: 10
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 83.33% Gaps: 0
DB: 14
US-09-851-138C-138 (1-12) x AY231584 (1-1794)
QY 1 LeuGluTyArgAsnAlaSerGlyLeuTyr 10
Db 827 CTGGAGTACAGGATGCTGCGCTATAT 856
RESULT 10
AY231585 LOCUS AY231585 1794 bp RNA linear VRL 01-MAR-2004
DEFINITION Hepatitis C virus isolate NB193 polyprotein gene, partial cds.
ACCESSION AY231585
VERSION AY231585.1 GI:37961931
KEYWORDS
SOURCE
ORGANISM
Hepatitis C virus
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE
1 (bases 1 to 1794)
Chaudhuri,S. and Naik,T.N.
Molecular epidemiology of HCV infection among acute and chronic
liver disease patients in Kolkata, India
Unpublished
JOURNAL
AUTHORS
2 (bases 1 to 1794)
Chaudhuri,S. and Naik,T.N.
Direct Submission
JOURNAL
TITLE
Submitted (07-FEB-2003) Division of Virology, National Institute of
Cholera and Enteric Diseases, P-33 CIT Road, Scheme- XM,
Beliaghata, Kolkata, WB 700010, India
FEATURES
Location/Qualifiers
source
1..1794
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="NB179"
/db_xref="taxon:11103"
/note="genotype: 3b"
254..>1794
/note="contains core and envelope proteins"
/contig_start=1
/product="polyprotein"
/protein_id="AAP69954.1"
/db_xref="GI:37961930"
/translacion="MSTLPKPKROTENTLRPKKVPKAGQIVGVVLPKPPQPL
GVREVCAYEWSQPRGRTPTKAPREGSWAQGYWPVLYGEGCGWAGWLLPPRG
SRPWSQNDPRRRNRNGKVIDTLCGFADLMGYIPLIGAPVGGVARALAHGVRLED
GVNATGNLPCGFSIFLLALFSLCTCPASSLEYRNAGSLYLLTNDGNSRISVYEAD
VILHPCGVCVETDNNNTSCWTPISPTVAVKHGVTAGIRNHNMLVAPPTLCSAL
YVEDAFGASVILGQAFTRPROHKTVOCTNCISIVPGHVSGRHMAWMMWSPALGLV
ISHLNRLPQTFDLVGAHWGMAGLAYFSMQGNWAKVIVILMFSGVDATHTHTGGS
AAQTAGTSTFTFRGSPQNLQVNSNGSWHNSLTALNCNDSLTGTAGLFIYHKFNS
SGCPRMSCKPITYFNQGWGLPTDANINGPSDRPYCMHYPPRPNCITKPLNARYI
ASLPAP"
ORIGIN
Alignment Scores: 0.101 Length: 1794
Pred. No.: 10.00 Matches: 10
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 83.33% Gaps: 0
DB: 14
US-09-851-138C-138 (1-12) x AY231584 (1-1794)
QY 1 LeuGluTyArgAsnAlaSerGlyLeuTyr 10
Db 827 CTGGAGTACAGGATGCTGCGCTATAT 856

```

/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="42"
/db_xref="taxon:11103"
/note="genotype: 3b"
254..>1799
/note="contains core and envelope proteins"
/codon_start=1
/product="polyprotein"
/protein_id="AAP69957.1"
/db_xref="GI:37961936"
/translation="MSTLPKPKQTENTLRPNKVPKPPAGQIVGEVYVLPKPPGQQL
SRPVGQNDPRRRSRNLGKVIDTLTCGFADLMGYIPLIGAPVGGVARALAHGVRLEDD
GVNATGNLPGCSFSIFLLALFSLCTCPASSLEYRNASGLYLLTNDCSNRSIVYEADD
VILHLPGCVPCVETDNNNTSCWTPISPTVAVKHPGVTASIRHNHVMMLVAPPTLCSAL
YVEDAFGAVSLVGQAFTFRPRQHKVTQCNCSLYPGHVSGRHMMNMMNSPAIGLIV
ISHLMRLPQTFFDLVVGAWHGMAGLAYFSMQGNWAKVIVILMFSGVDATHTTTGGS
AAQATAGTFTFFTRGPSQNLQVNSNGSWHINSTALNCNDSLNTGFIAGLFYTHKFN
SGCPRMSSCKPITYFNQGWGGLTDANINGPSDRPYCWHYPPPCNITKPLNVCPG
ASLPAP"

CDS

ORIGIN

Alignment Scores:
Pred. No.: 0.101 Length: 1794
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 83.33% Indels: 0
DB: 14 Gaps: 0

US-09-851-138C-138 (1-12) x AY231587 (1-1794)

Qy 1 LeuGlutYrArgAsnAlaSerGlyLeuTyr 10
|||||
Db 827 CTGGAGTACAGGATGCGTCTGGCCTATAT 856

RESULT 12
AY231586
LOCUS AY231586 1799 bp RNA linear VRL 01-MAR-2004
DEFINITION Hepatitis C virus isolate NB57 polyprotein gene, partial cds.
ACCESSION AY231586
VERSION AY231586.1 GI:37961933
KEYWORDS
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.

REFERENCE 1 (bases 1 to 1799)
AUTHORS Chaudhuri,S. and Naik,T.N.
TITLE Molecular epidemiology of HCV infection among acute and chronic liver disease patients in Kolkata, India
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1799)
AUTHORS Chaudhuri,S. and Naik,T.N.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2003) Division of Virology, National Institute of Cholera and Enteric Diseases, P-33 CIT Road, Scheme- XM, Beliaghata, Kolkata, WB 700010, India

FEATURES

source
1..1799
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="NB57"
/db_xref="taxon:11103"
/note="genotype: 3b"
254..>1799
/note="contains core and envelope proteins"
/codon_start=1
/product="polyprotein"
/protein_id="AAP69956.1"
/db_xref="GI:37961934"
/translation="MSTLPKPKQTENTLRPNKVPKPPAGQIVGEVYVLPKPPGQQL

CDS

GVREVCRAVEWSQPRRRQPTPKARPREGRSQAQPGYPWPVLYGNEGCGWAGWLLPPRG
SRPVGQNDPRRRSRNLGKVIDTLTCGFADLMGYIPLIGAPVGGVARALAHGVRLEDD
GVNATGNLPGCSFSIFLLALFSLCTCPASSLEYRNASGLYLLTNDCSNRSIVYEADD
VILHLPGCVPCVETDNNNTSCWTPISPTVAVKHPGVTASIRHNHVMMLVAPPTLCSAL
YVEDAFGAVSLVGQAFTFRPRQHKVTQCNCSLYPGHVSGRHMMNMMNSPAIGLIV
ISHLMRLPQTFFDLVVGAWHGMAGLAYFSMQGNWAKVIVILMFSGVDATHTTTGGS
AAQATAGTFTFFTRGPSQNLQVNSNGSWHINSTALNCNDSLNTGFIAGLFYTHKFN
SGCPRMSSCKPITYFNQGWGGLTDANINGPSDRPYCWHYPPPCNITKPLNVCPG
ILLHSQP"

ORIGIN

Alignment Scores:
Pred. No.: 0.101 Length: 1799
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 83.33% Indels: 0
DB: 14 Gaps: 0

US-09-851-138C-138 (1-12) x AY231586 (1-1799)

Qy 1 LeuGlutYrArgAsnAlaSerGlyLeuTyr 10
|||||
Db 827 CTGGAGTACAGGATGCGTCTGGCCTATAT 856

RESULT 13
AY231588
LOCUS AY231588 1799 bp RNA linear VRL 01-MAR-2004
DEFINITION Hepatitis C virus isolate NB134 polyprotein gene, partial cds.
ACCESSION AY231588
VERSION AY231588.1 GI:37961937
KEYWORDS
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.

REFERENCE 1 (bases 1 to 1799)
AUTHORS Chaudhuri,S. and Naik,T.N.
TITLE Molecular epidemiology of HCV infection among acute and chronic liver disease patients in Kolkata, India
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1799)
AUTHORS Chaudhuri,S. and Naik,T.N.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2003) Division of Virology, National Institute of Cholera and Enteric Diseases, P-33 CIT Road, Scheme- XM, Beliaghata, Kolkata, WB 700010, India

FEATURES

source
1..1799
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="NB134"
/db_xref="taxon:11103"
/note="genotype: 3b"
254..>1799
/note="contains core and envelope proteins"
/codon_start=1
/product="polyprotein"
/protein_id="AAP69958.1"
/db_xref="GI:37961938"
/translation="MSTLPKPKQTENTLRPNKVPKPPAGQIVGEVYVLPKPPGQQL

CDS

GVREVCRAVEWSQPRRRQPTPKARPREGRSQAQPGYPWPVLYGNEGCGWAGWLLPPRG
SRPVGQNDPRRRSRNLGKVIDTLTCGFADLMGYIPLIGAPVGGVARALAHGVRLEDD
GVNATGNLPGCSFSIFLLALFSLCTCPASSLEYRNASGLYLLTNDCSNRSIVYEADD
VILHLPGCVPCVETDNNNTSCWTPISPTVAVKHPGVTASIRHNHVMMLVAPPTLCSAL
YVEDAFGAVSLVGQAFTFRPRQHKVTQCNCSLYPGHVSGRHMMNMMNSPAIGLIV
ISHLMRLPQTFFDLVVGAWHGMAGLAYFSMQGNWAKVIVILMFSGVDATHTTTGGS
AAQATAGTFTFFTRGPSQNLQVNSNGSWHINSTALNCNDSLNTGFIAGLFYTHKFN
SGCPRMSSCKPITYFNQGWGGLTDANINGPSDRPYCWHYPPPCNITKPLNVCPG
ILLHSQP"

ORIGIN

Alignment Scores:

Pred. No.: 0.101 Length: 1799
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 83.33% Indels: 0
DB: 14 Gaps: 0

US-09-851-138C-138 (1-12) x AY231588 (1-1799)

Qy 1 LeuGlutTyArGAsnAlaSerGlyLeuTyr 10
Db 827 CTGGAGTACAGGAATGCGTCTGGCCTATAT 856

RESULT 14
AY231589
LOCUS Hepatitis C virus isolate NB211 RNA linear VRL 01-MAR-2004
DEFINITION Hepatitis C virus isolate NB211 polyprotein gene, partial cds.
ACCESSION AY231589
VERSION AY231589.1 GI:37961939
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS Chaudhuri, S. and Naik, T.N.
TITLE Molecular epidemiology of HCV infection among acute and chronic liver disease patients in Kolkata, India
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1799)
AUTHORS Chaudhuri, S. and Naik, T.N.
TITLE Direct Submission
JOURNAL

REFERENCE
AUTHORS Chaudhuri, S. and Naik, T.N.
TITLE Molecular epidemiology of HCV infection among acute and chronic liver disease patients in Kolkata, India
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1799)
AUTHORS Chaudhuri, S. and Naik, T.N.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2003) Division of Virology, National Institute of Cholera and Enteric Diseases, P-33 CIT Road, Scheme- XM, Beliaghata, Kolkata, WB 700010, India
FEATURES
source
1..1799
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="NB211"
/db_xref="taxon:11103"
/note="genotype: 3b"
254..>1799
/note="contains core and envelope proteins"
/codon_start=1
/product="polyprotein"
/protein_id="AAP69959.1"
/db_xref="GI:37961940"

ORIGIN
Alignment Scores:
Pred. No.: 0.101 Length: 1799
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 83.33% Indels: 0
DB: 14 Gaps: 0
US-09-851-138C-138 (1-12) x AY231589 (1-1799)
Qy 1 LeuGlutTyArGAsnAlaSerGlyLeuTyr 10
Db 827 CTGGAGTACAGGAATGCGTCTGGCCTATAT 856

RESULT 15
AY231590
LOCUS Hepatitis C virus isolate 236 polyprotein gene, partial cds.
DEFINITION Hepatitis C virus isolate 236 polyprotein gene, partial cds.
ACCESSION AY231590
VERSION AY231590.1 GI:37961941
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS Chaudhuri, S. and Naik, T.N.
TITLE Molecular epidemiology of HCV infection among acute and chronic liver disease patients in Kolkata, India
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1799)
AUTHORS Chaudhuri, S. and Naik, T.N.
TITLE Direct Submission
JOURNAL

REFERENCE
AUTHORS Chaudhuri, S. and Naik, T.N.
TITLE Molecular epidemiology of HCV infection among acute and chronic liver disease patients in Kolkata, India
JOURNAL Submitted (07-FEB-2003) Division of Virology, National Institute of Cholera and Enteric Diseases, P-33 CIT Road, Scheme- XM, Beliaghata, Kolkata, WB 700010, India
FEATURES
source
1..1799
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="236"
/db_xref="taxon:11103"
/note="genotype: 3b"
254..>1799
/note="contains core and envelope proteins"
/codon_start=1
/product="polyprotein"
/protein_id="AAP69950.1"
/db_xref="GI:37961942"

ORIGIN
Alignment Scores:
Pred. No.: 0.101 Length: 1799
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 83.33% Indels: 0
DB: 14 Gaps: 0
US-09-851-138C-138 (1-12) x AY231590 (1-1799)
Qy 1 LeuGlutTyArGAsnAlaSerGlyLeuTyr 10
Db 827 CTGGAGTACAGGAATGCGTCTGGCCTATAT 856

Search completed: March 3, 2005, 18:33:08
Job time : 472.169 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:41:12 ; Search time 390.974 Seconds
(without alignments)
1239.345 Million cell updates/sec

Title: US-09-851-138c-190

Perfect score: 10

Sequence: 1 VKSPCAATAS 10

Scoring table: ~~OMFCO~~
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9400332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlp
-O=/cgn2.1/USPTO.spool.p/US09851138/runat.28022005.120306.21465/app.query.fasta_1.1123
-DB=GenEmbl -QFWT=fastap -SUFFIX=olig.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09851138 @CGN 1.1.6331 @runat.28022005.120306.21465 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPRLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sv.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	100.0	447	6 A50396	A50396 Sequence 51
2	10	100.0	447	6 AR127536	AR127536 Sequence
3	10	100.0	447	14 HPCOREAL	L39317 Hepatitis C
4	8	80.0	977	1 STMRIION	M88615 Streptomyce

5	8	80.0	977	6	A23996
AK063402	8	80.0	1415	8	AK063402
BC019997	10	80.0	1956	10	BC019997
AK110120	8	80.0	2696	8	AK110120
AB044076	1	80.0	4896	1	AB044076
AE014768	1	80.0	12617	1	AE014768
AE012359	1	80.0	13043	1	AE012359
AC145333	2	80.0	91185	2	AC145333
AL603749	9	80.0	102313	9	AL603749
H3127820	9	80.0	135259	9	H3127820
AP005255	8	80.0	143908	8	AP005255
AP003448	8	80.0	149887	8	AP003448
AP005768	8	80.0	151054	8	AP005768
AC019262	2	80.0	159802	2	AC019262
AL451066	2	80.0	166258	2	AL451066
AP003214	8	80.0	179428	8	AP003214
AC009083	2	80.0	194548	2	AC009083
AC113107	10	80.0	207239	10	AC113107
AC120168	2	80.0	215391	2	AC120168
AE016782	1	80.0	300620	1	AE016782
AL391112	1	80.0	300800	1	AL391112
SC0939122	1	80.0	311000	1	SC0939122
AX492786	6	80.0	349980	6	AX492786
AX553953	6	80.0	349980	6	AX553953
AX047772	98	70.0	70.0	98	AX047772
AX997408	100	70.0	70.0	100	AX997408
BV182280	201	70.0	70.0	201	BV182280
AY657826	477	70.0	70.0	477	AY657826
AF439536	546	70.0	70.0	546	AF439536
AY737547	765	70.0	70.0	765	AY737547
AX433814	780	70.0	70.0	780	AX433814
AR394626	783	70.0	70.0	783	AR394626
AB190575	803	70.0	70.0	803	AB190575
CHKPROTAM	818	70.0	70.0	818	CHKPROTAM
AB014967	909	70.0	70.0	909	AB014967
AB014968	914	70.0	70.0	914	AB014968
BX935117	921	70.0	70.0	921	BX935117
CGL290426	939	70.0	70.0	939	CGL290426
AX659986	981	70.0	70.0	981	AX659986
BX936168	1077	70.0	70.0	1077	BX936168
AR352272	1155	70.0	70.0	1155	AR352272

ALIGNMENTS

RESULT 1	A50396	Sequence 51 from Patent WO9613590.	447 bp	DNA	linear	PAT 07-MAR-1997
LOCUS	A50396	Sequence 51 from Patent WO9613590.	447 bp	DNA	linear	PAT 07-MAR-1997
DEFINITION	A50396	Sequence 51 from Patent WO9613590.	447 bp	DNA	linear	PAT 07-MAR-1997
ACCESSION	A50396	Sequence 51 from Patent WO9613590.	447 bp	DNA	linear	PAT 07-MAR-1997
VERSION	A50396.1	GI:2303407				
KEYWORDS	unidentified					
SOURCE	unidentified					
ORGANISM	unclassified.					
REFERENCE	1 (bases 1 to 447)					
AUTHORS	Maertens,G. and Stuyver,L.					
TITLE	NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS					
JOURNAL	PATENT: WO 9613590-A 51 09-MAY-1996;					
COMMENT	INNOGENETICS NV (BE)					
FEATURES	Other publication AU 3844095 960523.					
source	Location/Qualifiers					
	1. 447					
	/organism="unidentified"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:32644"					
ORIGIN						
Alignment Scores:						
Pred. No.:	0.169					447
Score:	10.00					10
Percent Similarity:	100.00%					0

Best Local Similarity: 100.00%	Mismatches: 0
Query Match: 100.00%	Indels: 0
DB: 6	Gaps: 0
US-09-851-138C-190 (1-10) x A50396 (1-447)	
QY 1 VallySerProCysAlaAlaThrAlaSer 10	
DB 265 GTGAAGTCGCCCTGCGCGCCACCGCCTCT 294	
RESULT 2	
LOCUS ARI27536	
DEFINITION Sequence 51 from patent US 6180768.	447 bp DNA linear PAT 16-MAY-2001
ACCESSION ARI27536	
VERSION ARI27536.1 GI:14114131	
KEYWORDS	
SOURCE Unknown.	
ORGANISM Unknown.	
REFERENCE 1 (bases 1 to 447)	
AUTHORS Maertens,G. and Stuyver,L.	
TITLE Sequences of hepatitis C virus genotypes and their use as prophylactic, therapeutic and diagnostic agents	
JOURNAL Patent: US 6180768-A 51 30-JAN-2001;	
FEATURES	
source	Location/Qualifiers
1..447	
/organism="unknown"	
/mol_type="unassigned DNA"	
ORIGIN	
Alignment Scores:	
Pred. No.:	0.169 Length: 447
Score:	10.00 Matches: 10
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
DB: 6	Gaps: 0
US-09-851-138C-190 (1-10) x ARI27536 (1-447)	
QY 1 VallySerProCysAlaAlaThrAlaSer 10	
DB 265 GTGAAGTCGCCCTGCGCGCCACCGCCTCT 294	
RESULT 3	
HPCCOREAL	
LOCUS HPCCOREAL	
DEFINITION Hepatitis C virus type 3 clone NL96 precursor protein gene, partial cds.	447 bp ss-RNA linear VRL 16-OCT-2001
ACCESSION L39317	
VERSION L39317.1 GI:845497	
KEYWORDS Hepatitis C virus type 3	
SOURCE Hepatitis C virus type 3	
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.	
REFERENCE 1 (bases 1 to 447)	
AUTHORS van Doorn,L.J., Kletter,B., Stuyver,L., Maertens,G., Brouwer,H., Schalm,S., Heijlink,R. and Quint,W.	
TITLE Analysis of hepatitis C virus genotypes by a line probe assay and correlation with antibody profiles	
JOURNAL J. Hepatol. 21 (1), 122-129 (1994)	
MEDLINE 95052487	
PUBMED 7525693	
REFERENCE 2 (bases 1 to 447)	
AUTHORS van Doorn,L.J., Kletter,G.E., Stuyver,L., Maertens,G., Brouwer,J.T., Schalm,S.W., Heijlink,R.A. and Quint,W.G.	
TITLE Sequence analysis of hepatitis C virus genotypes 1 to 5 reveals multiple novel subtypes in the Benelux countries	
JOURNAL J. Gen. Virol. 76 (Pt 7), 1871-1876 (1995)	
MEDLINE 97201609	
PUBMED 9049395	
us-09-851-138c-190.olig.rge	
FEATURES	
source	Location/Qualifiers
1..447	
/organism="Hepatitis C virus type 3"	
/mol_type="genomic RNA"	
/db_xref="taxon:40363"	
/clone="NL96"	
/note="genotype: 3"	
<1..>447	
/codon_start=1	
/product="precursor protein"	
/protein_id="AA67840.1"	
/db_xref="GI:845498"	
/translation="DCINPATNLPQCSFSIFLLALFSLCLTPTAGLEARNASGLYMV TNDGNGSTVYBAGDIIILHPCVPCVRSGNTSRWCIVSPVTVAVKSPCAATASLRTH VDMVGAATLCSALYVGDLCCGALFLVGQGSFNRHQRHWTVQDCNCISI"	
mat_peptide	
1..96	
/product="core protein"	
/note="putative"	
97..447	
/product="el protein"	
/note="putative"	
ORIGIN	
Alignment Scores:	
Pred. No.:	0.169 Length: 447
Score:	10.00 Matches: 10
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
DB: 14	Gaps: 0
US-09-851-138C-190 (1-10) x HPCCOREAL (1-447)	
QY 1 VallySerProCysAlaAlaThrAlaSer 10	
DB 265 GTGAAGTCGCCCTGCGCGCCACCGCCTCT 294	
RESULT 4	
STMTRIBON	
LOCUS STMTRIBON	
DEFINITION Streptomyces aureofaciens ribonuclease gene, complete cds.	977 bp DNA linear BCT 26-APR-1993
ACCESSION M88615	
VERSION M88615.1 GI:153423	
KEYWORDS extracellular guanylspecific ribonuclease; ribonuclease gene.	
SOURCE Streptomyces aureofaciens	
ORGANISM Streptomyces aureofaciens	
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	
REFERENCE 1 (bases 1 to 977)	
AUTHORS Nazarov,V., Sevcik,J., Durcova,G. and Stanssens,P.	
JOURNAL Unpublished (1992)	
COMMENT Original source text: Streptomyces aureofaciens (strain R8/26) DNA.	
FEATURES	
source	Location/Qualifiers
1..977	
/organism="Streptomyces aureofaciens"	
/mol_type="genomic DNA"	
/strain="R8/26"	
/db_xref="taxon:1894"	
149..640	
/gene="ribonuclease"	
149..640	
/gene="ribonuclease"	
/EC_number="3.1.4.8"	
/codon_start=1	
/translation="MTRSKSPLVWGVAVILAVAGVYLLAGRGSGSTHFKAAASSAAA GTSPKSGAPKPSPPAGGAWTPADALADVCTKLPSQADTLALIAKNGPYPNRDLG VVFENRESRLPKKNGYHEFTVTPGSDNRTRRVVTGGYGEQYVSPDHYATFQELD PRC"	
sig_peptide	
149..349	


```

mat_peptide
/gene="ribonuclease"
350.637
/gene="ribonuclease"
/product="ribonuclease"
/EC_number="3.1.4.8"

ORIGIN
Alignment Scores:
Pred. No.: 36.1 Length: 977
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 1 Gaps: 0

US-09-851-138C-190 (1-10) x STMRIEON (1-977)

QY 3 SerProCyeAlaAlaThrAlaSer 10
Db 949 TCGCCATGCCGACGACGCGATCG 972

RESULT 5
LOCUS A23996 977 bp DNA linear PAT 18-JAN-1995
DEFINITION S.aureofaciens R08/26 sarnae.
ACCESSION A23996
VERSION A23996.1 GI:809619
KEYWORDS Streptomyces aureofaciens
ORGANISM Streptomyces aureofaciens
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE 1 (bases 1 to 977)
AUTHORS Nazarov V., Botterman J., Stanssens P. and Sevcik J.
TITLE A novel ribonuclease and its inhibitor
JOURNAL Patent: EP 0537399-A 1 21-APR-1993;
PLANT GENETIC SYSTEMS, N.V
FEATURES
Location/Qualifiers
1..977
/gene="Streptomyces aureofaciens"
/mol_type="unassigned DNA"
/db_xref="taxon:1894"

ORIGIN
Alignment Scores:
Pred. No.: 36.1 Length: 977
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-190 (1-10) x A23996 (1-977)

QY 3 SerProCyeAlaAlaThrAlaSer 10
Db 949 TCGCCATGCCGACGACGCGATCG 972

RESULT 6
LOCUS AK063402 1415 bp mRNA linear PLN 24-JUL-2003
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:001-115-A04, full
insert sequence.
ACCESSION AK063402
VERSION AK063402.1 GI:32973420
KEYWORDS FLI CDNA; oligo-capping.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1
AUTHORS The Rice Full-Length cDNA Consortium, National Institute of

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
2 (bases 1 to 1415)
Adachi J., Aizawa K., Akimura T., Arakawa T., Carninci P., Doi K.,
Fujimura T., Fukuda S., Hanagaki T., Hara A., Hashizume W.,
Hayashida K., Hayashizaki Y., Hayatsu N., Hiramoto K., Hiraoka T.,
Hori F., Hotta I., Iida Y., Ikeda R., Imamura K., Imoto K.,
Imotani K., Ishibiki J., Ishii Y., Ishikawa M., Itoh M., Kagawa I.,
Kanagawa S., Katoh H., Kawagashira N., Kawai J., Kawamata M.,
Kikuchi S., Kishikawa-Hirozane T., Kishimoto N., Kobayashi M.,
Kodama T., Kojima K., Kojima Y., Kondo S., Konno H., Kouda M.,
Koya S., Kurahara C., Kurotsaki T., Kusumegi T., Li C., Lu M.,
Masuda H., Matsubara K., Matsuyama T., Miura J., Miyazaki A.,
Mizuno K., Murakami K., Murata M., Nagata T., Nakamura M.,
Namiki T., Narikawa R., Niikura J., Nishi K., Nomura K.,
Numasaki R., Ohneda E., Ohno M., Ohtsuki K., Oka M., Ooka H.,
Osato N., Ota Y., Ootomo Y., Ryu R., Saitoh H., Sakai C., Sakai K.,
Sakazume N., Sano H., Sasaki D., Sato K., Satoh K., Shibata K.,
Shinagawa A., Shiraki T., Shishiki T., Sogabe Y., Sugano S.,
Sugiyama A., Suzuki K., Suzuki Y., Tagami M., Tagami-Takeda Y.,
Tagawa A., Takahashi F., Takaku-Akahira S., Tanaka T., Tomaru A.,
Tova T., Tsunoda Y., Ueda M., Waki K., Xie Q., Yahagi W.,
Yamada H., Yamamoto M., Yasunishi A., Yazaki J., Yokomizo S. and
Yoshimura A.
Direct Submission
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression: 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan [E-mail:skikuchi@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007]
This clone is one of the 28K full-length cDNA clones from japonica
rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi S., Satoh K.,
Nagata T., Kawagashira N., Doi K., Kishimoto N., Yazaki J.,
Ishikawa M., Yamada H., Ooka H., Hotta I., Kojima K., Namiki T.,
Ohneda S., Yahagi W., Suzuki K., Li C., Ohtsuki K., Shishiki T. and
Yamamoto M.
FAIS Genome Sequencing & Analysis Group: Ootomo Y., Iida Y.,
Fujimura T., Ikeda R., Ishibiki J., Kawamata M., Kobayashi M.,
Kodama T., Kurotsaki T., Kusumegi T., Lu M., Masuda H., Miura J.,
Mizuno K., Narikawa R., Niikura J., Oka M., Ryu R., Sugano S.,
Sugiyama A., Suzuki Y., Tsunoda Y., Ueda M., Xie Q., Yokomizo S.,
Yoshimura A., Matsubara K. and Murakami K.
Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi J., Aizawa K.,
Akimura T., Arakawa T., Carninci P., Fukuda S., Hanagaki T.,
Hara A., Hashizume W., Hayashida K., Hayatsu N., Hiramoto K.,
Hiraoka T., Hori F., Iida Y., Imamura K., Imotani K., Ishii Y.,
Itoh M., Kagawa I., Kanagawa S., Katoh H., Kawai J.,
Kishikawa-Hirozane T., Kojima Y., Kondo S., Konno H., Kouda M.,
Koya S., Kurahara C., Matsuyama T., Miyazaki A., Murata M.,
Nakamura M., Nishi K., Nomura K., Numasaki R., Ohno M., Ohta N.,
Ota Y., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,

```

Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toyota, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

FEATURES
source
1. .1415
Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="001-115-A04"

ORIGIN

Alignment Scores:
Pred. No.: 50.7 Length: 1415
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 8 Gaps: 0

US-09-851-138C-190 (1-10) x AK063402 (1-1415)

Qy 3 SerProCysAlaAlaThrAlaSer 10

Db 316 AGCCGTCGCGCGCCACTGCATCT 339

RESULT 7

BC019997
LOCUS
DEFINITION
Mus musculus RIKEN cDNA 0610038K03 gene, mRNA (cdna clone MGC:28693 IMAGE:4240949), complete cds.

ACCESSION
BC019997.1 GI:18044105

KEYWORDS
MGC.

SOURCE
Mus musculus (house mouse)

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1956)

REFERENCE
AUTHORS
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buotow, K.H., Schaefter, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettaman, M., Maman, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, M.A., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL
PUBMED
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932

REFERENCE
2 (bases 1 to 1956)

Strausberg, R.

Direct Submission

TITLE
Submitted (19-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
NTH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT
Contact: MGC help desk

Email: cgaps@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 37 Row: m Column: 21
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21312009.

FEATURES

source

Location/Qualifiers
1. .1956
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:28693 IMAGE:4240949"
/tissue_type="Kidney, normal. 5 month old male mouse."
/clone_lib="NCI CGAP_Kid14"
/lab_host="DH10B"
/note="Vector: pCMV-SpOrTe"

gene

1. .1956
/gene="0610038K03Rik"
/db_xref="LocusID:68371"
/db_xref="MGI:1915621"
2. .868
/gene="0610038K03Rik"
/codon_start=1
/product="RIKEN CDNA 0610038K03"
/protein_id="AAH19997.1"
/db_xref="GI:18044106"
/db_xref="LocusID:68371"
/db_xref="MGI:1915621"
/translation="MKLPFIADAFATAGNPAACVCLLETLERDAHQIOAREMNL SETAFRLKQPTDSFTQSSRFGLWFTPVSEVPLCGHATLASAALVPHKIQENSTLT FVTMGELKARRADEGVIDFPVYPTFFQDFHEVEDLIKAAIGDTLVQDIRYSTDTRK LTVRSLSDYDRSFLSKLNTPELPAIKETKVKRGLILTVKGEPGOTAPYDPYSRYF APWVGAEDPVTSGSAHTVLSSYWSQQLRKEMRAFCQSRRGGLDLSLRPDGRVDING GAVIVLEGTITA"

CDS

Alignment Scores:
Pred. No.: 68.3 Length: 1956
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 10 Gaps: 0

ORIGIN

US-09-851-138C-190 (1-10) x BC019997 (1-1956)
Qy 1 ValLysSerProCysAlaAlaThr 8
Db 222 GTGAAGTCCCTGTGCGGCACG 245

RESULT 8

LOCUS

AK110120

DEFINITION

Oryza sativa (japonica cultivar-group) cDNA clone:002-161-B06, full insert sequence.

ACCESSION

AK110120

VERSION

AK110120.1 GI:32995329

KEYWORDS

FLI CDNA; oligo capping.

SOURCE

Oryza sativa (japonica cultivar-group)

ORGANISM

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzeae; Oryza.

1
REFERENCE
AUTHORS
 The Rice Full-Length cDNA Consortium, National Institute of
 Agrobiological Sciences Rice Full-Length cDNA Project Team:
 Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
 Kiehimoto,N., Yabuchi,J., Iehikawa,M., Yamada,H., Ooka,H., Hotta,I.,
 Kojima,K., Namiki,T., Ohneda,E., Tahagi,W., Suzuki,K., Li,C.,
 Ohtsuka,K., Shishiki,T., Foundation of Advancement of International
 Science Genome Sequencing & Analysis Group: Ohtsuka,Y., Murakami,K.,
 Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
 Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
 Nariikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
 Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
 Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN,
 Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
 Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
 Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
 Saito,R., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T.,
 Yoshino,M., and Hayashizaki,Y.
 Collection, mapping, and annotation of over 28,000 cDNA clones from
 japonica rice
JOURNAL
MEDLINE
PUBMED
 Science 301 (5631), 376-379 (2003)
 12752273
 12859764

2 (bases 1 to 2696)
 Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
 Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
 Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
 Hori,F., Hotta,I., Iida,Y., Ikeda,R., Imamura,K.,
 Imotani,K., Ishibiki,J., Ishii,Y., Iehikawa,M., Itoh,M., Kagawa,I.,
 Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
 Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
 Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
 Koya,S., Kurihara,C., Kurotsaki,T., Kusumegi,T., Li,C., Lu,M.,
 Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
 Mizuno,K., Murakami,K., Murata,M., Negata,T., Nakamura,M.,
 Namiki,T., Nariikawa,R., Niikura,J., Nishi,K., Nomura,K.,
 Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
 Osato,N., Ota,Y., Ootomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K.,
 Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
 Shingawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
 Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
 Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
 Toyota,T., Teunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
 Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
 Yoshimura,A.
 Direct Submission
 Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
 Agrobiological Sciences, Department of Molecular Genetics, Head of
 Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
 Tel:81-29-838-7007, Fax:81-29-838-7007)
 This clone is one of the 28K full-length cDNA clones from japonica
 rice.
 URL : <http://cdna01.dna.affrc.go.jp/cDNA/>
 NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K.,
 Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
 Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T.,
 Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuka,K., Shishiki,T. and
 Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Ohtsuka,Y., Iida,Y.,
 Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M.,
 Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J.,
 Mizuno,K., Nariikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S.,
 Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S.,
 Yoshimura,A., Matsubara,K. and Murakami,K.
 Genome Exploration Research Group in Riken Genomic Sciences Center
 and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K.,
 Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T.,
 Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K.,
 Hirooka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y.,
 Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J.,
 Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,

Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
 Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E.,
 Ohta,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H.,
 Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T.,
 Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,
 Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toyota,T., Waki,K.,
 Yasunishi,A. and Hayashizaki,Y.
 Location/Qualifiers
 1..2696
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /culivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="002-161-B06"

ORIGIN

Alignment Scores:
 Pred. No.: 91.7 Length: 2696
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 80.00% Indels: 0
 DB: 8 Gaps: 0

US-09-851-138c-190 (1-10) x AK110120 (1-2696)
 QY 3 SerProCysAlaAlaThrAlaser 10
 DB 1259 TCACCGTGTGGCGGCGCGCGAGT 1282

RESULT 9
 AB044076/c
 LOCUS
 DEFINITION
 Myxococcus xanthus moka gene for hybrid sensor, complete cds.
 ACCESSION
 AB044076
 VERSION
 AB044076.1 GI:13516916
 KEYWORDS
 Myxococcus xanthus
 Myxococcus xanthus
 Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
 Cytophactariae; Myxococcaceae; Myxococcus.
 1 (sites)
 Kimura,Y., Nakano,H., Terasaka,H. and Takegawa,K.
 Myxococcus xanthus moka encodes a hybrid sensor required for
 development and osmotic tolerance
 Unpublished
 JOURNAL
 REFERENCE
 2 (bases 1 to 4896)
 Kimura,Y.
 Direct Submission
 Submitted (31-MAY-2000) Yoshio Kimura, Kagawa University, Faculty
 of Agriculture; 2393 Miki-cho Kagawa, Miki, Kagawa 761-0795, Japan
 (E-mail:kimura@ag.kagawa-u.ac.jp, Tel:81-87-891-3118,
 Fax:81-87-891-3021)
 Location/Qualifiers
 1..4896
 /organism="Myxococcus xanthus"
 /mol_type="genomic DNA"
 /db_xref="taxon:34"
 418..4572
 /gene="moka"
 /gene="moka"
 418..4572
 /codon_start=1
 /product="hybrid sensor"
 /protein_id="BAB4038.1"
 /db_xref="GI:13516917"

FEATURES
 source
 1..4896
 /organism="Myxococcus xanthus"
 /mol_type="genomic DNA"
 /db_xref="taxon:34"
 418..4572
 /gene="moka"
 /gene="moka"
 418..4572
 /codon_start=1
 /product="hybrid sensor"
 /protein_id="BAB4038.1"
 /db_xref="GI:13516917"

gene
 CDS

WVGTTGSLFFYENRTHFHVTEOGLPESVIPAMHEAHDGTLWVGTITGAVRYDQGRPT
RVSRAGQGLPDVDTAILVDGRDGTWLTQAGLARMGGGVTVYGPXKGFNTPIIVMVE
DSGRVWFGSDTGLVWDGQKGFQRFQDQGLPGDLVLALHADPDGTVMVGTFETGLG
RWEDTWARFTVAQGLVDDAVFSLVDPDGLWMSNKGVSRRRDLDDVAAGVRPR
LTTMDPTDRGMSAECNGTQPSANRGEGRLWFTNLRCALIVVDVPRVRESQPPEV
RIEVRVQGTFFPMGEFVELEPGASRIEIRFTAFTPVDAARLFFRRLAGHDDTWVHA
ESRRLYNGLRQPSYRFEVQAKRGDGTWTFVSLDVLPLEKLMQRTQFLLCVLQVGR
LGVSVLLRYGQDKRERWLAERVKORTQALANAELEANNHTLRQAQALVQVGR
AAGVQLAARGVIEHNNPLAVISNLEHASBEADALARELDGTRDVGARLDVNOALRD
ALLGADRVRIVODLKMFSRDPDEKQGPVELHVLDSAVKIAMGELRPRAKLVDYGD
VPRVEGNARLAQVFLNLLINAAQALPEGOAGTNEVRLVTRRPGDGRVTEVDTGSG
ISELLEGRIDFPFTFKPVGVTGGLSLCHAYLTAMGGTIAVDSRAGRSVPRVTLR
AAAEARVGGDGSVEPTASDRGVGAAPVPTASSGDIATVPDVSVEHGAASDAT
TMLRGEAAPRAELDSTPQASPTQAGWGTTPQAKSTDGRAADSAALAAQSQNG
GRWAAQVASTGMSRSEAGTVGSHQCEPSTHSAEGSPDORMGEGDRFSATEAT
RGVLVDDDALVSGAIRRTLARENDVDVLVSARRALEKLTGEARYDVVLCMLMPE
MTGMDLYDALAQVDSRRRSACSSPAVPRPPRGRSWSAWRIRAWRSPTKRCASS
GPKWRVLAGRHSSWPGRVGPAPAAWSGRVTRRRRACGHPRAPPRRRCGGRA
GHRRPV

ORIGIN

Alignment Scores:				
Pred. No.:	158	Length:	4896	
Score:	8.00	Matches:	8	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	80.00%	Indels:	0	
DB:	1	Gaps:	0	

US-09-851-138C-190 (1-10) x AB044076 (1-4896)

Qy 3 SerProCysAlaAlaThrAlaser 10

Db 2605 AGCCGTCGCTGCCAGCTTCG 2582

RESULT 10

AB014768

LOCUS

DEFINITION Bifidobacterium longum NCC2705 section 155 of 202 of the complete genome.

AB014768 AE014295

AB014768.1 GI:23326651

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Bifidobacterium longum NCC2705

Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;

Bifidobacteriaceae; Bifidobacterium.

1 (bases 1 to 12617)

Schell, M.A., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B.,

Pessi, G., Zwiolen, M.-C., Desiere, F., Bork, P., Delley, M.,

Pridmore, D. and Arigoni, F.

The genome sequence of Bifidobacterium longum reflects its

adaptation to the human gastrointestinal tract

Proc. Natl. Acad. Sci. U.S.A. 99 (22), 14422-14427 (2002)

12381787

2 (bases 1 to 12617)

Schell, M.A., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B.,

Pessi, G., Zwiolen, M.-C., Desiere, F., Bork, P., Delley, M.,

Pridmore, D. and Arigoni, F.

Direct Submission

Submitted (27-AUG-2002) Bioscience, Nestle Research Center, P. O.

Box 44, Lausanne 26 1000, Switzerland

Location/Qualifiers

1. 12617

/organism="Bifidobacterium longum NCC2705"

/mol_type="genomic DNA"

/strain="NCC2705"

/db_xref="taxon:206672"

188..2119

/locus tag="BL1379"

188..2119

/locus_tag="BL1379"

/note="COG family: predicted membrane GTPase involved in

stress response; PFAM_ID: GTP_EFTU; PFAM_ID: GTP_EFTU_D2
/codon_start=1
/transl_table=11
/product="widely conserved protein similar to those
annotated as GTP-binding elongation factor TypA/BiPa"
/protein_id="AA25180.1"
/db_xref="GI:23326652"
/translation="NAVRGDIRNVAIVAHVDHGHKTLTVNAMLQOQSHVFSEEREVPDRV
MSNDLREKGGITILAKNTAVEYTGPAAYKHGPEGITILNIIDTPGHADFGGSEVERGI
SVNDGVLLVDADEGLPQTRFVLRKALAKLPVLCVANKVDRPDARIEVCGETSDL
LGLADDVOHEGIDLDLQLLSMPVLYCAKAGYASVQPADGLPDNDNLPLEFTI
ISTIPAPTEGAPLOAHVANIDSSDFLGRGLVRYNGTLEKGYGLSRVDGSLN
FRVSELLRTQGLERIPVESAGPGDIVAVAGVDIMIGETIVDPNDPKPLIHVDDPA
IEMFTNDSPLAGTEGDKHKTARMIKDRLDRELIGNVSIKVLPTDRDAMVQGRG
ELALAVLAQMERREGVELTVRGPVVTIKIDGVEINPEMNTTIDVEYMGVTQJMA
DKRGMGNTNKGSGWVRVQFTVPSRGLIGFTALLSATRGTCGSSISAGYAPWAGA
ITIRQNGSVSRDKGIATPYAMQRLQARGNFVEPQSPVEGQGVGNKPKDLDVNI
TLAKHNNRSSTADVLETLTPTFKMSLESLEDFANEDECVEVTPESIRVRKILGRE
DNYKWRKQRQRAAQQGK"
2243..2665
/locus_tag="BL1380"
2243..2665
/locus_tag="BL1380"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA25181.1"
/db_xref="GI:23326653"
/translation="MRPWSYRONALVKSLITIAAGVASAFVCTFAHRMGAELNIPYGL
VIAFLILIGUSTCAKRMAGVGLALHLIASSTLAWGMAUTTTSGKALIVAGFGDMPP
FSQAGYIWLXGLILVQVLLILPARFWFIPTHSER"
2745..3722
/gene="pheA"
/locus_tag="BL1381"
2745..3722
/gene="pheA"
/locus_tag="BL1381"
/locus_tag="BL1381"
/note="PheA; COG family: prephenate dehydratase; PFAM_ID:
PPT; PFAM ID: ACT"
/codon_start=1
/transl_table=11
/product="probable chorismate mutase (CM); prephenate
dehydratase"
/protein_id="AA25182.1"
/db_xref="GI:23326654"
/translation="MARMLFYLGPBEGTTHOAMTAADQFATLGDPLIALPDVPAI
MQAVSRQGVIANENVEGVVNVNLDALIDAPNAAGFARISVDVAFVAFYRGHSI
YDCTGNPVAHPHGLAQCTRFIAEHLQPEPASAACRDLKPGRVALGPSICGLHY
DLDTAEHVODFDGAHTEFLVAPRDVQVELNARAHNSDAGDPETITIFLPTVPGVY
LADLLDLRDLAGLNMTSFSRPTKGHDGTYSFIATDAAPWEPFRFTALEEIAHGWDW
AKTLAVYPRRPNPVDAMMLPEGGVHITNRAPEGWQVTEARKELLW"
3716..4783
/gene="tyrA"
/locus_tag="BL1382"
3716..4783
/gene="tyrA"
/locus_tag="BL1382"
/note="TyrA; COG family: prephenate dehydratase;
PFAM_ID: PDH"
/codon_start=1
/transl_table=11
/product="probable prephenate dehydratase"
/protein_id="AA25183.1"
/db_xref="GI:23326655"
/translation="WVDTVGIVGLGIGSLARRLTERGVVVVNNRHPHPVAAEAD
GIFCKSTLSELMADPDDVVCNPLKMPALIAALPLMGDHPNTTLTDVSGVKGMVR
DVKQAGLGCYVGAHPHMGNSGWAADPHLYDGLWALTVDDESTDRFRFLDVAAM
ITKDVGNRVIVVDDETHDKAAAMI SHMPVSTALINELVANDPNRNIATAALAGVRD
MTRVSLTDPDTRAMVEEDSLNVEALLRRMAARLTDMDQLHAGAGEQDVMGFFAEG
DPFRYKAAVTHAGITAAQDGTATAAPQAGTTTAAGFPFERELAVPEAGWQQTLLFSARR
GEAITGFVHPQRAIIQLRPAM"
4955..5252
/locus_tag="BL1383"

Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Melandris, J.,
Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, J.M., Novo, M.T.M.,
Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
Rosi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,
Takita, M.A., Tamura, R.B., Teixeira, E.C., Tezza, R.I.D., Trindade dos
Santos, M., Truffi, D., Teal, S.M., White, F.F., Setubal, J.C. and
Kitajima, J.P.

TITLE

Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de
São Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
Brazil

FEATURES

source

1. 13043
/organism="Xanthomonas campestris pv. campestris str. ATCC
33913"
/mol_type="genomic DNA"
/strain="ATCC 33913"
/db_xref="ATCC:33913"
/db_xref="taxon:190485"
/note="pathovar: campestris"
complement(428. 1408)
/gene="XCC2470"
complement(428. 1408)
/gene="XCC2470"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="IS1479 transposase"
/protein_id="AA041746.1"
/db_xref="GI:21113631"
/translation="MOLTGDAEGLGRKQTRREIPLAEAEQVVPWQILGLVAPHYP
VSRGQEPYALATMLRIHLQQWYALSDPAMEALHEIPTLRFAQLGGLNDVPDET
TILNFRLLTHGLAARMLEAVNAHLARGQSURSTIVDATTIAAPSTKNTDARD
PEMRQTKNQWYGMKAHIGVDFSLGVHCHCTAANADVTVTHALHKGEDSVFG
DSYGTADKREELQCEAAFFIAAKRSVLQAIQNKREAREQREWEHFKAIVRAKVEHP
FRVYKQFGYKVRGLAKNTAQLVTLFALSNLWMKQQLPAMGSVRL"
complement(1730. 3346)
/gene="agla"
/note="synonym: XCC2471"
complement(1730. 3346)
/gene="agla"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="alpha-glucosidase"
/protein_id="AA041747.1"
/db_xref="GI:21113632"
/translation="MSQTPWRGAVIQIYIPRSFLDSNGDGLPGLIAKLDIAGL
GVDAIWISPEFKSPMADFGYDIADPAVDPLGSLVDPLLEKHAHGLGLKWMIDQVL
SHSSIAHWFOQRQSNPKADWYVADPDREDGTPPNWLSLFGVAMQWPREQY
YLHNFQDLNFAEYQQATLDNVFLRDGVDGFLDADINCFHDAQLRDNPAP
PADKRVGRFSADNPYAYQYFYFNTPENLAFRLRLGLDLYPNAVSLGISED
LATATRYQATTHMGYSFELLVQDYSAAYIRDTVSRLEATMLEGQPCWAI SNHDVVR
AVTRWGGAHATPAFARMVALLCSLRGSLCYQGEELGLSEAEVAFEDLQDPYGI.TFW
PTFKGRDGRCTPMPWTDPASAGTSGKPLPLAASHRAAVSVQDDAHSVLSVRDP
LAWRKEMPALRGSTAFYDTAEPLVLFREHAGQVULLAFNLSADPADLALPAGWEQ
VDVPGVELGAMGGHURLAGHAVVAAGRG"
3583. 4038
/gene="XCC2472"
3583. 4038
/gene="XCC2472"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="ISD1 transposase"
/protein_id="AA041748.1"

gene

CDS

gene
CDS

/db_xref="GI:21113633"
/translation="MHYGAWTSVTHCFDNRRFRSMVTVDHETHEWPDIVDQSLKGD
DVADAMTRIVAQRGKPTAKVDNGSEFSKVGSGVCIKNRVEPDFSGAANQITPFWK
ASTDTCGRSASTRVGSHWPTFKARSNNGDASITRSGGPTVWHGKLLKN"
4417. 4560
/gene="XCC2473"
4417. 4560
/gene="XCC2473"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="ISD1 transposase"
/protein_id="AA041749.1"
/db_xref="GI:21113634"
/translation="MQGRPMQMGFIERNFGSYPRGVLMNHIFSTLSEIRKQTEHLAD
YNQ"

gene

CDS

/gene="virB4"
/note="synonym: XCC2474"
complement(4707. 7160)
/gene="virB4"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="virB4 protein"
/protein_id="AA041750.1"
/db_xref="GI:21113635"
/translation="MFPDPSIFSEFVSISTHVPASVLTKTGGDYLLIWHLAGLPFVGR
DEFELEQRHNTFNRLVQTRAPANFAVHDVRRRKIGNSGRFKQSFNOALSTDEY
YAAASSQKIMQNELYFSMIRPVTKGLVEKSGDPDKIRVEEQAIAKVLELATNVE
AVLKDYSPVRLGMVEAKNGVVFSETLFLGYLLNRIIDEPVLOAPVAAYLVPVSKHMF
ANKTGDVFIRTPDGINHFGAILNVKEYADYTPGLNGKILDFYVTVHVSFSPVGRH
DALKLETKGMMISGDKSSQIRDLDMDDVASGNFVLGEYHFTVAVYSDSQEKL
ARQATTRAELSNAAGVFSSKEDLAIASSFYQLPQNWRPRTIRANLSLNLGLSPLH
NFAQOKHNPWGDAVTTLOTNGQYFYFNHATHPAENSLGEAKIGNTMVIGKSGTG
KTALINFLQSVQKQPDVPPTIFFPKQRGARIFVRACGNYLALENGSPTFGNPQCE
SNVNTQFLAELIKVLGKTRYSSREEDIYRACGNYLALENGSPTFGNPQCE
DDGLVRLRRYTAGNSLGVFPDVTIDITLRAIIGDYTDVINDVVRVPIYNYLL
HRELDIDGRFLIYMWDEFKWLIDEGGKLEFANQKTIIRKQGLGIFATQSPEDAL
ASDIAALIEQTATWLLPNPNASRDDYIDGLKLTDAEYQVVVLSDRSCLFLVKQGH
ASAVQNLNRGMDALSVSSISTDNIETIMHQLSRKAGKLGVSQGLTPEQWLEDFYA
NRRGSKRKSVKDKEVEDV"
complement(7268. 7579)
/gene="XCC2475"
complement(7288. 7579)
/gene="XCC2475"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="virB3 protein"
/protein_id="AA041751.1"
/db_xref="GI:21113636"
/translation="MHKDVLRGCTRPAMFLGVPIYPIFFIGAGGGLLGMVFNLYWILL
TPIVILFMQVQAKRDEMIFRLGLRLFRMRVNIQRYSGMWVFSFNEYRKHGAKR
Q"

gene

CDS

complement(7572. 7976)
/gene="virB2"
/note="synonym: XCC2476"
complement(7572. 7976)
/gene="virB2"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="virB2 protein"
/protein_id="AA041752.1"
/db_xref="GI:21113637"
/translation="MKPEIDKLAADAKMIGAKALKALVFTSLLLAGGAVATETGLG
DTGEKMGCFLLNNVGNLLMGSALVTIATVAGTQIAFAHRISEVSPILIGGVLLGA
AAQIANNVIGDGTQDCSTSSMMLQSAQIYA"

gene

CDS

complement(7572. 7976)
/gene="virB2"
/note="synonym: XCC2476"
complement(7572. 7976)
/gene="virB2"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="virB2 protein"
/protein_id="AA041752.1"
/db_xref="GI:21113637"
/translation="MKPEIDKLAADAKMIGAKALKALVFTSLLLAGGAVATETGLG
DTGEKMGCFLLNNVGNLLMGSALVTIATVAGTQIAFAHRISEVSPILIGGVLLGA
AAQIANNVIGDGTQDCSTSSMMLQSAQIYA"

```

gene      complement(8039..8872)
/genes="virB1"
/notes="synonym: XCC2477"
CDS       complement(8039..8872)
/genes="virB1"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="VirB1 protein"
/db_xref="GI:21113638"
/translation="MLPGLMGCTGLAVGEVQHVVRSSRPYALGVVGRILAR
EPRLFEAVATAMLEQYNSLGLQVNRNLYKYLTSYEMAFKCPNIVAASKI
LAECRSGANWKFSCYISNGFETGPKHGVQKIYASIROSVASGSEPIAVLPSR
PIKVRVENPLHTAAELADAVARRIDLSRGRTGNSQQGYMLAPPPLEQLPLQVVS

Alignment Scores:
Pred. No.:      389      Length:      13043
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      80.00%      Indels:      0
DB:              1          Gaps:      0

US-09-851-138C-190 (1-10) x AB012359 (1-13043)

Qy      3 SerProCyeAlaAlaThrAlaSer 10
|||||
Db      3551 TCGCCGTGCGCTGCGACTGCGTGG 3528
|||||

RESULT 12
AC145333
LOCUS      91185 bp      DNA      linear      HTG 23-JAN-2004
DEFINITION      Pan troglodytes clone rp43-22b18, WORKING DRAFT SEQUENCE, 26
unordered pieces.
ACCESSION      AC145333
VERSION      AC145333.2      GI:41151915
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE      Yu, J., Do, T. and Roe, B.A.
AUTHORS      Yu, J., Do, T. and Roe, B.A.
TITLE      Pan troglodytes BAC Clone rp43-22b18
JOURNAL      Unpublished
AUTHORS      Yu, J., Do, T. and Roe, B.A.
TITLE      Direct Submission
JOURNAL      Submitted (25-JUN-2003) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE      3 (bases 1 to 91185)
AUTHORS      Yu, J., Do, T. and Roe, B.A.
TITLE      Direct Submission
JOURNAL      Submitted (23-JAN-2004) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT      On Jan 23, 2004 this sequence version replaced gi:32189494.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code: UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

*      1      2446: contig of 2446 bp in length
*      2447 2546: gap of unknown length
*      2547 2547: contig of 2113 bp in length
*      2548 4659: gap of unknown length
*      4660 4759: gap of unknown length
*      4760 7221: contig of 2462 bp in length
*      7222 7321: gap of unknown length
*      7322 9926: contig of 2605 bp in length
*      9927 10026: gap of unknown length
*      10027 12160: contig of 2134 bp in length
*      12161 12260: gap of unknown length
*      12261 14942: contig of 2682 bp in length
*      14943 15042: gap of unknown length
*      15043 17221: contig of 2179 bp in length
*      17222 17321: gap of unknown length
*      17322 19644: contig of 2323 bp in length
*      19645 19744: gap of unknown length
*      19745 21954: contig of 2210 bp in length
*      21955 22054: gap of unknown length
*      22055 24715: contig of 2661 bp in length
*      24716 24815: gap of unknown length
*      24816 28005: contig of 3190 bp in length
*      28006 28105: gap of unknown length
*      28106 30608: contig of 2503 bp in length
*      30609 30708: gap of unknown length
*      30709 33707: contig of 2999 bp in length
*      33708 33807: gap of unknown length
*      33808 37154: contig of 3347 bp in length
*      37155 37254: gap of unknown length
*      37255 41030: contig of 3776 bp in length
*      41031 41130: gap of unknown length
*      41131 43718: contig of 2588 bp in length
*      43719 43818: gap of unknown length
*      43819 46987: contig of 3169 bp in length
*      46988 47087: gap of unknown length
*      47088 50024: contig of 2937 bp in length
*      50025 50124: gap of unknown length
*      50125 53593: contig of 3469 bp in length
*      53594 53693: gap of unknown length
*      53694 57248: contig of 3555 bp in length
*      57249 57348: gap of unknown length
*      57349 62165: contig of 4817 bp in length
*      62166 62265: gap of unknown length
*      62266 68192: contig of 5927 bp in length
*      68193 68292: gap of unknown length
*      68293 72992: contig of 4700 bp in length
*      72993 73092: gap of unknown length
*      73093 78528: contig of 5536 bp in length
*      78529 78728: gap of unknown length
*      78729 83439: contig of 4711 bp in length
*      83440 91185: gap of unknown length
*      91185 91185: contig of 7646 bp in length.

FEATURES
Location/Qualifiers
1..91185
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="rp43-22b18"
/clone_lib="RPCI - 43 Male Chimpanzee BAC Library"

ORIGIN

Alignment Scores:
Pred. No.:      2.32e+03      Length:      91185
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      80.00%      Indels:      0
DB:              2          Gaps:      0

US-09-851-138C-190 (1-10) x AC145333 (1-91185)

Qy      1 VallySerProCyeAlaAlaThr 8
|||||
Db      89786 GTTAAATCCCATGTGCAGCTACT 89809

```



```

RESULT 13
AL603749/c
LOCUS      102313 bp      DNA      linear      PRI 15-NOV-2001
DEFINITION Human DNA sequence from clone RP11-133N1 on chromosome 1, complete
            sequence.
ACCESSION  AL603749
VERSION     AL603749.6  GI:16973163
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 102313)
AUTHORS     Whitehead, S.
TITLE       Direct Submission
JOURNAL     Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
            On Nov 16, 2001 this sequence version replaced gi:15723828.
COMMENT     During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em., EMBL; Sw.,
            SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep
            This sequence was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-133N1 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-133N1. It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-133N1 is at 1 in this sequence. The
            true left end of clone RP4-597A16 is at 100314 in this sequence.
            The true right end of clone RP11-584P2 is at 85607 in this
            sequence.

FEATURES             Location/Qualifiers
     source           1..102313
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-133N1"
                     /clone_lib="RPCI-11.1"
misc_feature         55651..55661
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only"

ORIGIN
Alignment Scores:
Pred. No.:         2.58e+03      Length:      102313
Score:             8.00         Matches:      8
Percent Similarity: 100.00%     Conservative: 0
Best Local Similarity: 100.00%  Mismatches:    0
Query Match:       80.00%       Indels:        0
DB:                9           Gaps:         0

```

```

US-09-851-138C-190 (1-10) x AL603749 (1-102313)

Qy      3 SerProCysAlaAlaThrAlaSer 10
        |||||
Db      74126 TCTCCCTGTGCAGCTACTGCTCA 74103

RESULT 14
HS127B20 135259 bp      DNA      linear      PRI 05-JUN-2003
LOCUS     Human DNA sequence from clone RP1-127B20 on chromosome 22 Contains
DEFINITION the 3' end of the ARHGAP8 gene for Rho GTPase activating protein 8,
            an RPL6 (60S Ribosomal protein L6) pseudogene, the gene for a novel
            PHD finger protein, ESTe, STSs, GSSs, genomic marker D22S274 and a
            ca repeat polymorphism, complete sequence.
            283838
ACCESSION 283838.2 GI:6572184
KEYWORDS  HTG; 60S ribosomal protein L6; ARHGAP8; ca repeat polymorphism;
           D22S274; GTPase activating protein; PHD finger; RPL6.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE   1 (bases 1 to 135259)
AUTHORS     Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Hunt, A.
TITLE       Direct Submission
JOURNAL     Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
            On Dec 13, 1999 this sequence version replaced gi:2276307.
COMMENT     During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep -----
            Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -----
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            RP1-127B20 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCVPAC2
            This sequence is the entire insert of clone RP1-127B20 The true
            right end of clone CTA-116F5 is at 36125 in this sequence.
            Location/Qualifiers
            1..135259
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="RZPD:RPCI704B20127"
            /db_xref="taxon:9606"
            /chromosome="22"
            /clone="RP1-127B20"

```



```
misc_feature
1. .330
/clone_lib="RPCI-1"
/note="match: STS: Em:AL008674"
123. .48171
/gene="ARHGAP8"
mRNA
join(<123. .209,7822. .7920,10934. .11044,30720. .30793,
33424. .33501,34383. .34511,45182. .45285,47726. .48171)
/gene="ARHGAP8"
/product="dJ127B20.1 (Rho GTPase activating protein 8)"
/note="match: Em:U62794 Em:U02570 Em:223024
match: ESTs: Em:W52941 Em:AA553320 Em:A1635875 Em:H55220
Em:AA508035 Em:A1808551 Em:AA916728 Em:A1380836
Em:A1825280 Em:A1346478 Em:AA857992 Em:AA535608
Em:AA658030 Em:A1743003 Em:AA536082 Em:AA514271
Em:A1280634 Em:A1299314 Em:AA657736 Em:A1284954 Em:H55390"
/evidence="not experimental"
CDS
join(<123. .209,7822. .7920,10934. .11044,30720. .30793,
33424. .33501,34383. .34511,45182. .45285,47726. .48046)
/gene="ARHGAP8"
/note="match: proteins: Sw:Q07960 Tr:O14988 Tr:O15376
Wp:CE03768"
/codon_start=2
/evidence="not experimental"
/product="dJ127B20.1 (Rho GTPase activating protein 8)"
/protein_id="CA662993.1"
/db_xref="GI:6572195"
/translation="YKKNLKALYVVHPTSFIRKVLNLIKPLSHKFGKKVIYFNLYSE
LHEHLKYDQIVTPPEVRYDEKLQSLHEGRTPPTPPRPLPQFGVSLQYLKD
KNQGLIPVLATFTVRLREKGLREGLPFRSASVQTVREIQLYNQKGPVNFDDYGD
IHIPAVILKTFURELPQPLLTFQAYEQILGTCVSSLRVATGCRQLLSLPHNVVL
RYLMGLFHAVRESIFKNMNSNLAFCVGLNIWPSQGVSSIALVPLNMFELLLIEY
YKSTFTPEAGHEGLAPWEOGSRAAPLQEAIVPTQATGLTKPTLPSPLMVAARRLL"
complement(476. .1058)
/note="MER77 repeat: matches 1. .589 of consensus"
repeat_region
1302. .1315
/note="2.8 copies 5 mer GGGGT 28% conserved"
repeat_region
1356. .1717
/gene="ARHGAP8"
/note="match: STS: Em:G65368"
1628. .1637
/note="3.3 copies 3 mer GCT 20% conserved"
1768. .1777
/note="2.5 copies 4 mer CCAAT 20% conserved"
1887. .1906
/note="2.9 copies 7 mer CCCTTGT 31% conserved"
1890. .1910
/note="2.6 copies 8 mer TTGTCCT 33% conserved"
1982. .2278
/note="AluJo repeat: matches 1. .306 of consensus"
2303. .2323
/note="5.2 copies 4 mer CTTT 26% conserved"
2304. .2320
/note="3.4 copies 5 mer TTCT 27% conserved"
complement(2324. .2637)
/note="AluSg repeat: matches 1. .313 of consensus"
2301. .2316
/note="2.3 copies 7 mer ATGTGAC 32% conserved"
2973. .3273
/note="AluSg repeat: matches 1. .298 of consensus"
3284. .3561
/note="AluJb repeat: matches 1. .279 of consensus"
3618. .3915
/note="AluSp repeat: matches 1. .297 of consensus"
3959. .4128
/note="MIR repeat: matches 34. .213 of consensus"
complement(4136. .4428)
/note="AluJo repeat: matches 22. .312 of consensus"
complement(4510. .4817)
/note="AluJb repeat: matches 1. .310 of consensus"
4841. .5142
/note="AluSg repeat: matches 1. .301 of consensus"
complement(5207. .5344)
/note="MIR repeat: matches 34. .182 of consensus"
```

```
repeat_region
5374. .5532
/note="MIR3 repeat: matches 8. .174 of consensus"
5666. .5676
/note="2.2 copies 5 mer GAGAG 22% conserved"
6074. .6083
/note="2.5 copies 4 mer CCTC 20% conserved"
6301. .6322
/note="3.7 copies 6 mer CTGCGC 35% conserved"
6324. .6353
/note="2.5 copies 12 mer CTGAGCTGTGGG 51% conserved"
6502. .6516
/note="3.0 copies 5 mer CATAA 21% conserved"
6601. .6891
/note="AluSg repeat: matches 1. .292 of consensus"
7747. .7758
/note="3.0 copies 4 mer CCCT 24% conserved"
complement(7825. .8293)
/note="match: GSS: Em:B14191"
complement(7899. .8318)
/note="match: STS: Em:AL021650"
8100. .8110
/note="2.2 copies 5 mer GAAAT 22% conserved"
complement(8386. .8491)
/note="MIR3 repeat: matches 15. .136 of consensus"
complement(8605. .8936)
/note="3.7 copies 6 mer GGGCTT 35% conserved"
9360. .9381
/note="2.6 copies 5 mer GGCAG 26% conserved"
9583. .9595
/note="2.6 copies 5 mer TCTTT 26% conserved"
9600. .9698
/note="MIR repeat: matches 50. .152 of consensus"
9968. .10020
/note="L2 repeat: matches 3235. .3299 of consensus"
10059. .10071
/note="2.6 copies 5 mer TCTTT 26% conserved"
complement(10072. .10378)
/note="AluSx repeat: matches 6. .312 of consensus"
complement(10476. .10570)
/note="L2 repeat: matches 3215. .3308 of consensus"
10771. .10794
/note="12.0 copies 2 mer TG 30% conserved"
10998. .11015
/note="2.0 copies 9 mer CCGCGGCC 27% conserved"
11202. .11371
/note="5.2 copies 33 mer GCTCGCTGCTCGGTGCTGTGGGGGCGCGT
277% conserved"
11552. .11564
/note="2.2 copies 6 mer TAGGTG 26% conserved"
11693. .11807
/note="MLT11 repeat: matches 297. .449 of consensus"
complement(11897. .12176)
/note="AluSx repeat: matches 1. .279 of consensus"
12188. .12206
/note="1.9 copies 10 mer ATTTTGAAC 38% conserved"
12347. .12356
/note="2.5 copies 4 mer TTTC 20% conserved"
12391. .12787
/note="MLT1D repeat: matches 4. .449 of consensus"
complement(12788. .13093)
/note="AluSx repeat: matches 1. .308 of consensus"
13094. .13152
/note="MLT1D repeat: matches 449. .501 of consensus"
```

Alignment Scores:

Pred. No.:	3.33e+03	Length:	135259
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	80.00%	Indels:	0
DB:	9	Gaps:	0

US-09-851-138C-190 (1-10) x HS127B20 (1-135259)

[illegible]

Db 74177 GTGAATCGCCCTGCGCGCCACC 74200

Search completed: March 3, 2005, 18:35:43
Job time : 439.974 secs

```

/codon_start=1
/product="hypothetical protein"
/protein_id="BAD31236.1"
/db_xref="GI:50509538"
/translation="NGADPPAIDGAGEVGEVSKIRBEGGKNGKGGNHRREYFP
FLYCVGRREARAGFAAAALGHGGSGGRTGTTGGPHPSARVAGGPARQRARGE
RPSRPRGEGEGREGRGRRKKRRGRRLWPRAGRGGGLWAKLGPKEKED
YFLLFIFNKL"
complement(21496..22438)
/gene="OSJNBb0087F05.7"
/mRNA
complement(join(<21496..21644,22378..>22438))
/gene="OSJNBb0087F05.7"
/notes="start and end point are not identified"
complement(join(21496..21644,22378..22438))
/gene="OSJNBb0087F05.7"
/notes="predicted by GlimmerM etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD31237.1"
/db_xref="GI:50509539"
/translation="MWKYGSAQQRPQLNPPYSTRAEGRGARTGTGVTGTPRPHGPRW
TGRTRAHTCRHRGSHASAAPEGAGG"
23800..24699
/gene="OSJNBb0087F05.8"
/misc_feature
23800..24699
/gene="OSJNBb0087F05.8"
/notes="probably inactive due to including stop codon(s) in
CDS
pseudogene, retrotransposon RIRE2, orf3"
/pseudo
24789..36531
/gene="OSJNBb0087F05.9"
/misc_feature
24789..36531
/gene="OSJNBb0087F05.9"
/notes="probably inactive due to including stop codon(s) in
CDS
pseudogene, retrotransposon RIRE2, orf4"
/pseudo
37373..37822
/gene="OSJNBb0087F05.10"
/misc_feature
37373..37822
/gene="OSJNBb0087F05.10"
/notes="probably inactive due to no initiation codon in CDS
pseudogene, retrotransposon RIRE2, orf5"
/pseudo
complement(38187..41255)
/gene="OSJNBb0087F05.11"
complement(38187..41255)
/gene="OSJNBb0087F05.11"
/notes="probably inactive due to including frameshift(s) in
CDS
pseudogene, retrotransposon RIRE2, orf2 GAG-POL precursor"
/pseudo
complement(41322..44293)
/gene="OSJNBb0087F05.12"
complement(41322..44293)
/gene="OSJNBb0087F05.12"
/notes="probably inactive due to including stop codon(s) in
CDS
pseudogene, retrotransposon RIRE2, orf1 GAG-POL precursor"
/pseudo

```

Alignment Scores:

Pred. No.:	3.53e+03	Length:	143908
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	80.00%	Indels:	0
DB:	8	Gaps:	0

US-09-851-138C-190 (1-10) x AP005255 (1-143908)

QY 1 VallyserProCysAlaAlaThr 8

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:41:12 ; Search time 508.267 Seconds

(without alignments)
1239.345 Million cell updates/sec

Title: US-09-851-138c-174
Perfect score: 13
Sequence: 1 VRSGNTSRCWIPV 13

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9404695

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters: -DB=frame+ p2n.model -DEV=xlp
-Q/cnq2_1/USPTO spool_p/US09851138/runat_28022005 120306 21465/app_query.fasta_1.1123
-DB=GenEmbl -QFMT=fastap -SURFIX=oligo -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09851138 @CGN 1 1 6331 @runat_28022005 120306 21465 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.ste.*
12: gb.tv.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	100.0	447	6 A50396	A50396 Sequence 51
2	13	100.0	447	6 ARI27536	ARI27536 Sequence
3	13	100.0	447	14 HPCCOREAL	L39317 Hepatitis C
4	11	84.6	1584	14 HPCJK072A9	D49753 Hepatitis C

5	10	76.9	181	14	S64511	S64511 (envelope r
6	10	76.9	277	6	AR066642	AR066642 Sequence
7	10	76.9	306	6	AR066617	AR066617 Sequence
8	10	76.9	333	6	AR066621	AR066621 Sequence
9	10	76.9	333	6	AR066631	AR066631 Sequence
10	10	76.9	357	14	AF515919	AF515919 Hepatitis
11	10	76.9	357	14	AF515922	AF515922 Hepatitis
12	10	76.9	384	14	HCUI4196	U4196 Hepatitis C
13	10	76.9	384	14	HCUI4201	U4201 Hepatitis C
14	10	76.9	411	14	HPCCP3	D30046 Hepatitis C
15	10	76.9	576	6	AR035884	AR035884 Sequence
16	10	76.9	576	6	AR035885	AR035885 Sequence
17	10	76.9	576	6	AR035887	AR035887 Sequence
18	10	76.9	576	6	I20120	I20120 Sequence 26
19	10	76.9	576	6	I20121	I20121 Sequence 27
20	10	76.9	576	6	I20123	I20123 Sequence 29
21	10	76.9	576	6	AR340298	AR340298 Sequence
22	10	76.9	576	6	AR340299	AR340299 Sequence
23	10	76.9	576	6	AR340301	AR340301 Sequence
24	10	76.9	576	14	HPCBP1ET	L16673 Hepatitis C
25	10	76.9	576	14	HPCBP1EU	L16674 Hepatitis C
26	10	76.9	576	14	HPCBP1W	L16650 Hepatitis C
27	10	76.9	872	14	AB107935	AB107935 Hepatitis
28	10	76.9	872	14	AB107936	AB107936 Hepatitis
29	10	76.9	875	14	AB107947	AB107947 Hepatitis
30	10	76.9	1195	14	HPCOPRP	L38334 Hepatitis C
31	10	76.9	1195	14	HPCOPRQ	L38335 Hepatitis C
32	10	76.9	1195	14	HPCOPRR	L38336 Hepatitis C
33	10	76.9	1584	14	HPCJK128B3	D49756 Hepatitis C
34	10	76.9	1584	14	HPVJK070A8	D49752 Hepatitis C
35	10	76.9	2551	6	E04262	E04262 cDNA encodi
36	10	76.9	2551	6	E04807	E04807 cDNA to 5'-
37	10	76.9	9416	14	AF238482	AF238482 Hepatitis
38	10	76.9	9416	14	AF238483	AF238483 Hepatitis
39	10	76.9	9416	14	AF238484	AF238484 Hepatitis
40	10	76.9	9416	14	AF238485	AF238485 Hepatitis
41	10	76.9	9589	6	E07361	E07361 GRNA of Hep
42	10	76.9	9589	6	E07362	E07362 cDNA of Hep
43	10	76.9	9589	6	I12861	I12861 Sequence 2
44	10	76.9	9589	14	HPCPOLP	D00944 Hepatitis C
45	10	76.9	9674	14	AB047640	AB047640 Hepatitis

ALIGNMENTS

RESULT 1	A50396	Sequence 51 from Patent WO9613590.	447 bp	DNA	linear	PAT 07-MAR-1997
A50396	A50396	A50396.1	GI:2303407			
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						
FEATURES						
source						
1..447						
/organism="unidentified"						
/mol_type="unassigned DNA"						
/db_xref="taxon:32644"						

Alignment Scores:	4.21e-06	Length:	447
Pred. No.:	13.00	Matches:	13
Score:	100.00%	Conservative:	0
Percent Similarity:			

```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-174 (1-13) x A50396 (1-447)
Qy 1 ValArgSerGlyAenThrSerArgCysTrpIleProVal 13
Db 211 GTACGCTCTGCAATACATCAAGATGCTGGATCCCTGTG 249

RESULT 2
LOCUS AR127536 linear PAT 16-MAY-2001
DEFINITION Sequence 51 from patent US 6180768.
ACCESSION AR127536
VERSION AR127536.1 GI:14114131
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 447)
AUTHORS Maertens,G. and Stuyver,L.
TITLE Sequences of hepatitis C virus genotypes and their use as prophylactic, therapeutic and diagnostic agents
JOURNAL Patent: US 6180768-A 51 30-JAN-2001;
FEATURES
source Location/Qualifiers
1..447
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.: 4.21e-06 Length: 447
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-174 (1-13) x AR127536 (1-447)
Qy 1 ValArgSerGlyAenThrSerArgCysTrpIleProVal 13
Db 211 GTACGCTCTGCAATACATCAAGATGCTGGATCCCTGTG 249

RESULT 3
HPCOREAL
LOCUS HPCOREAL linear VRL 16-OCT-2001
DEFINITION Hepatitis C virus type 3 clone NL96 precursor protein gene, partial cds.
ACCESSION L39317
VERSION L39317.1 GI:845497
KEYWORDS
SOURCE Hepatitis C virus type 3
ORGANISM Hepatitis C virus type 3
REFERENCE 1 (bases 1 to 447)
AUTHORS van Doorn,L.J., Kleter,B., Stuyver,L., Maertens,G., Brouwer,H., Schalm,S., Heijtkink,R. and Quint,W.
TITLE Analysis of hepatitis C virus genotypes by a line probe assay and correlation with antibody profiles
JOURNAL J. Hepatol. 21 (1), 122-129 (1994)
MEDLINE 95052487
PUBMED 7525693
REFERENCE 2 (bases 1 to 447)
AUTHORS van Doorn,L.J., Kleter,G.E., Stuyver,L., Maertens,G., Brouwer,J.T., Schalm,S.W., Heijtkink,R.A. and Quint,W.G.
TITLE Sequence analysis of hepatitis C virus genotypes 1 to 5 reveals multiple novel subtypes in the Benelux countries
JOURNAL J. Gen. Virol. 76 (Pt 7), 1871-1876 (1995)
MEDLINE 97201609
PUBMED 9049395

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-174 (1-13) x A50396 (1-447)
Qy 1 ValArgSerGlyAenThrSerArgCysTrpIleProVal 13
Db 211 GTACGCTCTGCAATACATCAAGATGCTGGATCCCTGTG 249

RESULT 2
LOCUS AR127536 linear PAT 16-MAY-2001
DEFINITION Sequence 51 from patent US 6180768.
ACCESSION AR127536
VERSION AR127536.1 GI:14114131
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 447)
AUTHORS Maertens,G. and Stuyver,L.
TITLE Sequences of hepatitis C virus genotypes and their use as prophylactic, therapeutic and diagnostic agents
JOURNAL Patent: US 6180768-A 51 30-JAN-2001;
FEATURES
source Location/Qualifiers
1..447
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.: 4.21e-06 Length: 447
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-174 (1-13) x AR127536 (1-447)
Qy 1 ValArgSerGlyAenThrSerArgCysTrpIleProVal 13
Db 211 GTACGCTCTGCAATACATCAAGATGCTGGATCCCTGTG 249

RESULT 3
HPCOREAL
LOCUS HPCOREAL linear VRL 10-FEB-1999
DEFINITION Hepatitis C virus isolate JK072 gene for core, env, and part of E2/NS1, partial cds.
ACCESSION D49753
VERSION D49753.1 GI:1197124
KEYWORDS core, env, and part of E2/NS1.
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
REFERENCE 1 (sites)
AUTHORS Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tsuda,F., Lesmana,L.A., Miyakawa,Y. and Mayumi,M.
TITLE Hepatitis C virus variants from Jakarta, Indonesia classifiable into novel genotypes in the second (2e and 2f), tenth (10a) and eleventh (11a) genetic groups
JOURNAL J. Gen. Virol. 77 (Pt 2), 293-301 (1996)
MEDLINE 96226020
PUBMED 8627233
REFERENCE 2 (bases 1 to 1584)
AUTHORS Okamoto,H.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1584)
AUTHORS Okamoto,H.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1995) Hiroaki Okamoto, Jichi Medical School, Immunology Division; Minamikawachi-machi, Kawachi-gun, Tochigi 329-04, Japan (E-mail:hokamoto@jichi.ac.jp, Tel:0285-44-2111(ex.3334), Fax:0285-44-1557)
FEATURES
source Location/Qualifiers
1..1584
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="JK072"
```

```
/db_xref=taxon:11103"
<1..339
/citation=[2]
/evidence=not_experimental
340..>1584
/citation=[2]
/codon_start=1
/evidence=not_experimental
/product="core, env, and part of E2/NS1"
/protein_id="BAA08587.1"
/db_xref="GI:1197125"
/translation="MSTLPKPKQRTKNTNRPPQDKPPGGQIVGGVVVLPRRGPKL
GVRVKTSESRQPSRQPIPRARTEGRSWAQGYWPPLYGNEGCWAGWLLSPRG
SRPWPNDPRRRNLGKVIDLTTCGFADLMGYIPLAGAPVGGVARALAGVRALED
GINFATGNLPCCSFSLALSCLLTPTAGLEYRNSGLYMTVNDNSNIVYEATD
IILHPGVCPCVRSNGTSRCWISISPTVAVSPGAATASLRTHVDMVMYGAATLCSALY
IGDLCAFLVGOGFSWRHROLWTQECNSIYPGHLTGHAMDMNMMNSPAMTMVY
SQVRLPOTILDVLGAHWGMAGVAYYSMQGNWAKFVLVLCFLPSGVDAHTTTTGGVA
ARTTSGLTALFTTGGKQHIQLIN"
ORIGIN
Alignment Scores:
Pred. No.: 0.00307 Length: 1584
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 84.62% Indels: 0
DB: 14 Gaps: 0
US-09-851-138C-174 (1-13) x HPCJK072A9 (1-1584)
Qy 1 ValArgSerGlyAanThrSerArgCysTrpIle 11
Db 1027 GTACGCTCTGGCAACACATCAAGATGCTGGATC 1059
RESULT 5
S64511
LOCUS 181 bp mRNA linear VRL 30-SEP-1993
DEFINITION {envelope region} [hepatitis C virus, type III, Japanese isolate,
mRNA Partial, 181 nt].
ACCESSION S64511
VERSION S64511.1 GI:404414
KEYWORDS Hepatitis C virus
Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 181)
AUTHORS Kao,J.H., Chen,P.J., Lai,M.Y. and Chen,D.S.
TITLE Superinfection of heterologous hepatitis C virus in a patient with
chronic type C hepatitis
JOURNAL Gastroenterology 105 (2), 583-587 (1993)
MEDLINE 93328063
PUBMED 8392958
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 136193] from the original journal article.
FEATURES
source
1..181
/organism="Hepatitis C virus"
/mol_type="mRNA"
/db_xref=taxon:11103"
/note="type: III"
ORIGIN
Alignment Scores:
Pred. No.: 0.00571 Length: 181
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 14 Gaps: 0
US-09-851-138C-174 (1-13) x S64511 (1-181)
```

```
Qy 4 GlyAanThrSerArgCysTrpIleProVal 13
Db 128 GGGAAATACATCTCGTGCTGGATACCGGTC 157
RESULT 6
AR066642
LOCUS 277 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 42 from patent US 5851759.
ACCESSION AR066642
VERSION AR066642.1 GI:5997864
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 277)
AUTHORS Weiner,A.J.
TITLE Heteroduplex tracking assay (HTA) for genotyping HCV
JOURNAL Patent: US 5851759-A 42 22-DEC-1998;
FEATURES Location/Qualifiers
source
1..277
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Alignment Scores:
Pred. No.: 0.00857 Length: 277
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 6 Gaps: 0
US-09-851-138C-174 (1-13) x AR066642 (1-277)
Qy 4 GlyAanThrSerArgCysTrpIleProVal 13
Db 140 GGGAAATACATCTCGTGCTGGATACCGGTC 169
RESULT 7
AR066617
LOCUS 306 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 17 from patent US 5851759.
ACCESSION AR066617
VERSION AR066617.1 GI:5997839
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 306)
AUTHORS Weiner,A.J.
TITLE Heteroduplex tracking assay (HTA) for genotyping HCV
JOURNAL Patent: US 5851759-A 17 22-DEC-1998;
FEATURES Location/Qualifiers
source
1..306
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Alignment Scores:
Pred. No.: 0.00943 Length: 306
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 6 Gaps: 0
US-09-851-138C-174 (1-13) x AR066617 (1-306)
Qy 4 GlyAanThrSerArgCysTrpIleProVal 13
Db 73 GGGAAATACATCTCGTGCTGGATACCGGTC 102
```

```

RESULT 8
LOCUS       AR066621                      333 bp    DNA    linear    PAT 29-SEP-1999
DEFINITION   Sequence 21 from patent US 5851759.
ACCESSION   AR066621
VERSION     AR066621.1  GI:5997843
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 333)
AUTHORS    Weiner,A.J.
TITLE      Heteroduplex tracking assay (HTA) for genotyping HCV
JOURNAL    Patent: US 5851759-A 21 22-DEC-1998;
FEATURES    Location/Qualifiers
             source
               1..333
               /organism="unknown"
               /mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.:      0.0102      Length:      333
Score:          10.00      Matches:     10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:    76.92%      Indels:      0
DB:             6          Gaps:          0

US-09-851-138C-174 (1-13) x AR066621 (1-333)

Qy      4  GlyAsnThrSerArgCysTrpIleProVal 13
      |||||||
Db      115  GGAATACATCTCGGTGCTGGATACCGGTC 144

RESULT 9
LOCUS       AR066631                      333 bp    DNA    linear    PAT 29-SEP-1999
DEFINITION   Sequence 31 from patent US 5851759.
ACCESSION   AR066631
VERSION     AR066631.1  GI:5997853
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 333)
AUTHORS    Weiner,A.J.
TITLE      Heteroduplex tracking assay (HTA) for genotyping HCV
JOURNAL    Patent: US 5851759-A 31 22-DEC-1998;
FEATURES    Location/Qualifiers
             source
               1..333
               /organism="unknown"
               /mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.:      0.0102      Length:      333
Score:          10.00      Matches:     10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:    76.92%      Indels:      0
DB:             6          Gaps:          0

US-09-851-138C-174 (1-13) x AR066631 (1-333)

Qy      4  GlyAsnThrSerArgCysTrpIleProVal 13
      |||||||
Db      115  GGAATACATCTCGGTGCTGGATACCGGTC 144

RESULT 10
LOCUS       AF515919                      357 bp    RNA    linear    VRL 29-MAR-2003
DEFINITION   Hepatitis C virus isolate MRS39 envelope protein E1 (E1) gene,
             partial cds.
ACCESSION   AF515919
VERSION     AF515919.1  GI:29365672
KEYWORDS    .
SOURCE      Hepatitis C virus
ORGANISM    Hepatitis C virus
REFERENCE   1 (bases 1 to 357)
AUTHORS    Cantaloube,J.F., Biagini,P., Attoui,H., Gallian,P., de Micco,P. and
             de Lamballerie,X.
TITLE      Evolution of hepatitis C virus in blood donors and their respective
             recipients
JOURNAL    J. Gen. Virol. 84 (Pt 2), 441-446 (2003)

```

```

ACCESSION   AF515919
VERSION     AF515919.1  GI:29365666
KEYWORDS    .
SOURCE      Hepatitis C virus
ORGANISM    Hepatitis C virus
REFERENCE   1 (bases 1 to 357)
AUTHORS    Cantaloube,J.F., Biagini,P., Attoui,H., Gallian,P., de Micco,P. and
             de Lamballerie,X.
TITLE      Evolution of hepatitis C virus in blood donors and their respective
             recipients
JOURNAL    J. Gen. Virol. 84 (Pt 2), 441-446 (2003)
MEDLINE    22447295
PUBMED     12560577
REFERENCE   2 (bases 1 to 357)
AUTHORS    Cantaloube,J.F., de Micco,P. and de Lamballerie,X.
TITLE      Direct Submission
JOURNAL    Submitted (29-MAY-2002) Emerging Viruses Department, EFS
             Alpes-Mediterranee, 149 Boulevard Baillie, Marseille 13005, France
             Location/Qualifiers
             source
               1..357
               /organism="Hepatitis C virus"
               /virion
               /mol_type="genomic RNA"
               /isolate="MRS39"
               /db_xref="taxon:11103"
               /note="genotype: 2a"
               <1..>357
               /gene="E1"
               <1..>357
               /gene="E1"
               /codon_start=1
               /product="envelope protein E1"
               /protein_id="AA083226.1"
               /db_xref="GI:29365667"
               /translation="ERVGNTSRCWIPVSPNVAVQRPGLTQGLRAHIDIVVMSATLCS
             ALYVGDLCGVMLAAQMFTVSPQHHWVQECNSIYPGAISGRHAWDMWMMNSPTAT
             MILAYAMRIPEVIIDII"

ORIGIN
Alignment Scores:
Pred. No.:      0.0109      Length:      357
Score:          10.00      Matches:     10
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:    76.92%      Indels:      0
DB:             14          Gaps:          0

US-09-851-138C-174 (1-13) x AF515919 (1-357)

Qy      4  GlyAsnThrSerArgCysTrpIleProVal 13
      |||||||
Db      10  GGAACACATCTCGGTGTTGGATACCGGTC 39

RESULT 11
LOCUS       AF515922                      357 bp    RNA    linear    VRL 29-MAR-2003
DEFINITION   Hepatitis C virus isolate MRS44 envelope protein E1 (E1) gene,
             partial cds.
ACCESSION   AF515922
VERSION     AF515922.1  GI:29365672
KEYWORDS    .
SOURCE      Hepatitis C virus
ORGANISM    Hepatitis C virus
REFERENCE   1 (bases 1 to 357)
AUTHORS    Cantaloube,J.F., Biagini,P., Attoui,H., Gallian,P., de Micco,P. and
             de Lamballerie,X.
TITLE      Evolution of hepatitis C virus in blood donors and their respective
             recipients
JOURNAL    J. Gen. Virol. 84 (Pt 2), 441-446 (2003)

```



```
MEDLINE 22447295
PUBMED 12560577
REFERENCE 2 (bases 1 to 357)
AUTHORS Cantaloube,J.F., de Micco,P. and de Lamballerie,X.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2002) Emerging Viruses Department, EFS
Alpes-Mediterranee, 149 Boulevard Baille, Marseille 13005, France
FEATURES
source
1..357
/organism="Hepatitis C virus"
/virion
/mol_type="genomic RNA"
/isolate="WRS44"
/db_xref="taxon:11103"
/note="genotype: 2a"
<1..>357
/gene="E1"
<1..>357
/gene="E1"
/codon_start=1
/product="envelope protein E1"
/protein_id="AAO83229.1"
/db_xref="GI:29365673"
/translation="EKMGNTSRCHWIPSPNVAORPGALTQGLRAHIDMVMSATLCS
ADYGLCGVMLAQMFIIVSPQHHFVQECNCSIYPGTITGHRVANDMMWNSPTAT
MILAYMRVPEVIIDII"
ORIGIN
Alignment Scores:
Pred. No.: 0.0109 Length: 357
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 14 Gaps: 0
US-09-851-138C-174 (1-13) x AF515922 (1-357)
QY 4 GlyAenThSerArgCysTrpIleProVal 13
DB 10 GGAATACATCTCGGTGCTGGATACCGGTC 39
RESULT 12
HCUI4196
LOCUS HCUI4196 384 bp RNA linear VRL 27-JAN-1995
DEFINITION Hepatitis C virus 2a I31 envelope protein (ei) gene, partial cds.
ACCESSION U14196
VERSION U14196.1 GI:537645
KEYWORDS Hepatitis C virus
SOURCE Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 384)
AUTHORS Greene,W.K., Cheong,M.K., Ng,V. and Yap,K.W.
TITLE Prevalence of hepatitis C virus sequence variants in South-East
Asia
JOURNAL J. Gen. Virol. 76 (Pt 1), 211-215 (1995)
MEDLINE 95146953
PUBMED 7844535
REFERENCE 2 (bases 1 to 384)
AUTHORS Greene,W.K.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1994) Wayne K. Greene, Scitech Genetics, 12
Science Park Drive #04-04, Singapore 0511, Republic of Singapore
FEATURES
source
1..384
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/strain="2a"
/isolate="K43"
/specific_host="Homo sapiens"
/db_xref="taxon:11103"
/gene="E1"
<1..>384
/codon_start=1
/product="envelope protein"
/protein_id="AAC53890.1"
/db_xref="GI:537656"
/translation="AEVNTSRSRVMTNDCSNESITWQLQAAVLHVPGCVPCKEVGNT
SRCWIPSPNVAORPGALTQGLRTHIDMVMSATLCSALYVGDLCGGVMLAAQMFIV
SPQHHFVQECNCSIYPGTITGHRMA"
ORIGIN
Alignment Scores:
Pred. No.: 0.0117 Length: 384
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 14 Gaps: 0
US-09-851-138C-174 (1-13) x HCUI4196 (1-384)
QY 4 GlyAenThSerArgCysTrpIleProVal 13
DB 124 GGAATACATCTCGGTGCTGGATACCGGTT 153
RESULT 13
HCUI4201
LOCUS HCUI4201 384 bp RNA linear VRL 27-JAN-1995
DEFINITION Hepatitis C virus 2a K43 envelope protein (ei) gene, partial cds.
ACCESSION U14201
VERSION U14201.1 GI:537655
KEYWORDS Hepatitis C virus
SOURCE Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 384)
AUTHORS Greene,W.K., Cheong,M.K., Ng,V. and Yap,K.W.
TITLE Prevalence of hepatitis C virus sequence variants in South-East
Asia
JOURNAL J. Gen. Virol. 76 (Pt 1), 211-215 (1995)
MEDLINE 95146953
PUBMED 7844535
REFERENCE 2 (bases 1 to 384)
AUTHORS Greene,W.K.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1994) Wayne K. Greene, Scitech Genetics, 12
Science Park Drive #04-04, Singapore 0511, Republic of Singapore
FEATURES
source
1..384
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/strain="2a"
/isolate="K43"
/specific_host="Homo sapiens"
/db_xref="taxon:11103"
/gene="E1"
<1..>384
/codon_start=1
/product="envelope protein"
/protein_id="AAC53890.1"
/db_xref="GI:537656"
/translation="AEVNTSRSRVMTNDCSNESITWQLQAAVLHVPGCVPCKEVGNT
SRCWIPSPNVAORPGALTQGLRTHIDMVMSATLCSALYVGDLCGGVMLAAQMFIV
SPQHHFVQECNCSIYPGTITGHRMA"
ORIGIN
Alignment Scores:
Pred. No.: 0.0117 Length: 384
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 14 Gaps: 0
US-09-851-138C-174 (1-13) x HCUI4196 (1-384)
QY 4 GlyAenThSerArgCysTrpIleProVal 13
DB 124 GGAATACATCTCGGTGCTGGATACCGGTT 153
```

```
gene 1..384
CDS <1..>384
/gene="E1"
/codon_start=1
/product="envelope protein"
/protein_id="AAC53885.1"
/db_xref="GI:537646"
/translation="AQVKNISTGYMTNDCSNDSITWQLQAAVLHVPGCIPCEKVGNT
SRCWIPSPNVAORPGALTQGLRTHIDMVMSATLCSALYVGDLCGGVMLAAQMFII
SPRHHWFVQECNCSIYPGTITGHRMA"
ORIGIN
Alignment Scores:
Pred. No.: 0.0117 Length: 384
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 14 Gaps: 0
US-09-851-138C-174 (1-13) x HCUI4196 (1-384)
QY 4 GlyAenThSerArgCysTrpIleProVal 13
DB 124 GGAATACATCTCGGTGCTGGATACCGGTT 153
RESULT 13
HCUI4201
LOCUS HCUI4201 384 bp RNA linear VRL 27-JAN-1995
DEFINITION Hepatitis C virus 2a K43 envelope protein (ei) gene, partial cds.
ACCESSION U14201
VERSION U14201.1 GI:537655
KEYWORDS Hepatitis C virus
SOURCE Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 384)
AUTHORS Greene,W.K., Cheong,M.K., Ng,V. and Yap,K.W.
TITLE Prevalence of hepatitis C virus sequence variants in South-East
Asia
JOURNAL J. Gen. Virol. 76 (Pt 1), 211-215 (1995)
MEDLINE 95146953
PUBMED 7844535
REFERENCE 2 (bases 1 to 384)
AUTHORS Greene,W.K.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1994) Wayne K. Greene, Scitech Genetics, 12
Science Park Drive #04-04, Singapore 0511, Republic of Singapore
FEATURES
source
1..384
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/strain="2a"
/isolate="K43"
/specific_host="Homo sapiens"
/db_xref="taxon:11103"
/gene="E1"
<1..>384
/codon_start=1
/product="envelope protein"
/protein_id="AAC53890.1"
/db_xref="GI:537656"
/translation="AEVNTSRSRVMTNDCSNESITWQLQAAVLHVPGCVPCKEVGNT
SRCWIPSPNVAORPGALTQGLRTHIDMVMSATLCSALYVGDLCGGVMLAAQMFIV
SPQHHWFVQECNCSIYPGTITGHRMA"
ORIGIN
Alignment Scores:
Pred. No.: 0.0117 Length: 384
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 14 Gaps: 0
US-09-851-138C-174 (1-13) x HCUI4196 (1-384)
QY 4 GlyAenThSerArgCysTrpIleProVal 13
DB 124 GGAATACATCTCGGTGCTGGATACCGGTT 153
```

Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 14 Gaps: 0

US-09-851-138C-174 (1-13) x HCU14201 (1-384)

QY 4 GlyAsnThrSerArgCysTrrPileProVal 13
|||||
Db 124 GGAACACATCTCGGTGCTGGATACAGTC 153

RESULT 14

HPCCP3

LOCUS HPCCP3 411 bp RNA linear VRL 07-FEB-1999
DEFINITION Hepatitis C virus (individual isolate Td-3/93) gene for polypeptide precursor, partial cds (core protein (carboxy terminus) and E1 envelope protein (amino terminus half)).

ACCESSION

VERSION D30046.1 GI:485798

KEYWORDS E1 envelope protein; core protein.

SOURCE

ORGANISM Hepatitis C virus
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

REFERENCE

1 (sites)

Hotta,H., Handajani,R., Lusida,M.I., Soemarto,W., Doi,H.,

Miyajima,H. and Homma,M.

NS5b region analysis of hepatitis C virus in Indonesia on the basis of

J. Clin. Microbiol. 32 (12), 3049-3051 (1994)

JOURNAL

MEDLINE 95189942

PUBMED 7883898

REFERENCE

2 (bases 1 to 411)

Hotta,H.

Unpublished

REFERENCE 3 (bases 1 to 411)

Hotta,H.

Direct Submission

Submitted (28-Apr-1994) Hak Hotta, Kobe University School of

Medicine, Department of Microbiology; 7-5-1 Kusunoki-cho, Chuo-ku,

Kobe, Hyogo 650, Japan (Tel:078-341-7451(ex.3301),

Fax:078-351-6347)

Submitted (28-Apr-1994) to DDBJ by:

Hak Hotta

Kobe University School of Medicine

Department of Microbiology

7-5-1 Kusunoki-cho, Chuo-ku

Kobe, Hyogo 650

Japan

Phone: 078-341-7451 x3301

Fax: 078-351-6347.

FEATURES

source

1. .411 Location/Qualifiers

/organism="Hepatitis C virus"

/mol_type="genomic RNA"

/isolate="Td-3/93"

/db_xref="taxon:11103"

<1..>411

/note="The carboxy terminus of the core protein and the

amino terminal half of the E1 envelope protein of

hepatitis C virus"

/codon_start=1

/product="polypeptide precursor"

/protein_id="BA006282.1"

/db_xref="GI:485798"

/translation="LSCLTPTAGLEYNASGLYIVTDCNSSIIVYEAQDIILHMPG

CDPCVRSGNTSCWTVSVTVAGRAAATASLRTHVDMVGAATLCSALYVGLCGA

LFLVGGQFSWRHQHTVQDCNCISYPGHLTGHRM"

ORIGIN

Alignment Scores:

Pred. No.: 0.0125

Length: 411

Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 14 Gaps: 0

US-09-851-138C-174 (1-13) x HPCCP3 (1-411)

QY 1 ValArgSerGlyAsnThrSerArgCysTrrP 10
|||||
Db 145 GTACGCTCTGGCAACACATCAAGGTGCTGG 174

RESULT 15

AR035884

LOCUS AR035884 576 bp DNA linear PAT 29-SEP-1999

DEFINITION Sequence 26 from patent US 5871962.

ACCESSION AR035884

VERSION AR035884.1 GI:5952552

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 576)

AUTHORS Bukh,J., Miller,R.H. and Purcell,R.H.

TITLE Nucleotide and deduced amino acid sequences of the envelope 1 gene

of 51 isolates of hepatitis C virus and the use of reagents derived

from these sequences in diagnostic methods

Patent: US 5871962-A 26 16-FEB-1999;

Location/Qualifiers

1. .576

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:

Pred. No.: 0.0172

Score: 10.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 76.92%

DB: 6

Length: 576

Matches: 10

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-09-851-138C-174 (1-13) x AR035884 (1-576)

QY 4 GlyAsnThrSerArgCysTrrPileProVal 13
|||||
Db 124 GGAATACATCCGATGCTGGATACCGGTC 153

Search completed: March 3, 2005, 18:34:54

Job time : 510.267 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:41:12 ; Search time 430.072 Seconds

(without alignments)
1239.345 Million cell updates/sec

Title: US-09-851-138C-155

Perfect score: 11

Sequence: 1 VYEAGDIHLHL 11

Scoring table:

QAGAP 60.0 , XGAP 60.0
YAGAP 60.0 , YGAP 60.0
FAGAP 6.0 , FGAP 7.0
DEL 6.0 , DEL 7.0

Searched: 4708233 seqs, 24227607955 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9402261

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlp
-Q/cgn2 1/USPTO spool p/US09851138/runat 28022005 21465/app query.fasta 1.1123
-DB=GenEmbl -QFMT=fastp -SUFFIX=olig.rge -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09851138 @CGN 1 1 6331 @runat 28022005 120306 21465 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAP=60 -YGAPOP=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	100.0	447	6 A50396	A50396 Sequence 51
2	11	100.0	447	6 AR127536	AR127536 Sequence
3	11	100.0	447	14 HPCCKOREAL	L39317 Hepatitis C
4	11	100.0	474	14 AY739423	AY739423 Hepatitis

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	11	100.0	1584	14 HPCJK030A3	D49747 Hepatitis C
6	11	100.0	1584	14 HPVJK070A8	D49752 Hepatitis C
c 7	8	72.7	141167	9 AC093263	AC093263 Homo sapi
c 8	8	72.7	152290	2 AC016107	AC016107 Homo sapi
c 9	8	72.7	158618	2 AC139271	AC139271 Homo sapi
c 10	8	72.7	179668	2 AC139812	AC139812 Homo sapi
c 11	8	72.7	189938	2 AC119064	AC119064 Sus scrofa
c 12	8	72.7	189043	2 AC139804	AC139804 Homo sapi
c 13	8	72.7	196706	2 AC139462	AC139462 Homo sapi
c 14	8	72.7	198361	2 AC139814	AC139814 Homo sapi
c 15	7	63.6	446	6 CQ482459	CQ482459 Sequence
c 16	7	63.6	489	6 CQ503573	CQ503573 Sequence
c 17	7	63.6	489	6 CQ512403	CQ512403 Sequence
c 18	7	63.6	491	11 G49493	G49493 SHGC-68671
c 19	7	63.6	553	11 BV024710	BV024710 S209P6128
c 20	7	63.6	609	11 BV020094	BV020094 S212P6915
c 21	7	63.6	892	1 AF349070	AF349070 Unculture
c 22	7	63.6	990	8 BT008728	BT008728 Arabidops
c 23	7	63.6	990	8 AF401300	AF401300 Arabidops
c 24	7	63.6	1446	8 AY093053	AY093053 Arabidops
c 25	7	63.6	1504	8 AY087356	AY087356 Arabidops
c 26	7	63.6	2232	9 BC057848	BC057848 Homo sapi
c 27	7	63.6	2521	9 AK125305	AK125305 Homo sapi
c 28	7	63.6	2544	1 TTHTRSYN	M64273 T.thermophi
c 29	7	63.6	5708	1 AB091385	AB091385 Pseudomon
c 30	7	63.6	11007	5 CHKASMA	M13756 Gallus gall
c 31	7	63.6	14301	6 CQ572609	CQ572609 Sequence
c 32	7	63.6	16843	3 AC107205	AC107205 Leishmani
c 33	7	63.6	24552	2 AC017225	AC017225 Drosophila
c 34	7	63.6	39502	1 AY534910	AY534910 Unculture
c 35	7	63.6	4863	9 AL592161	AL592161 Human DNA
c 36	7	63.6	45550	9 AC006292	AC006292 Homo sapi
c 37	7	63.6	76875	2 AC018419	AC018419 Homo sapi
c 38	7	63.6	79355	2 AC021280	AC021280 Homo sapi
c 39	7	63.6	85217	8 ATT17J13	AL138651 Arabidops
c 40	7	63.6	93432	9 CNS01DSR	AL121840 Human chr
c 41	7	63.6	97911	2 AC091511	AC091511 Leishmani
c 42	7	63.6	99790	3 AC103910	AC103910 Leishmani
c 43	7	63.6	105306	8 ATP9D34	AL137081 Arabidops
c 44	7	63.6	110000	1 AE017261.12	Continuation (13 o
c 45	7	63.6	110000	2 BX511197_0	BX511197 Danio rer

ALIGNMENTS

RESULT 1	A50396	Sequence 51 from Patent WO9613590.	447 bp	DNA	linear	PAT 07-MAR-1997
LOCUS	A50396	Sequence 51 from Patent WO9613590.				
DEFINITION	A50396					
ACCESSION	A50396					
VERSION	A50396.1	GI:2303407				
KEYWORDS						
SOURCE	unidentified					
ORGANISM	unidentified					
REFERENCE	1 (bases 1 to 447)					
AUTHORS	Maertens, G. and Stuyver, L.					
TITLE	NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS					
JOURNAL	Patent: WO 9613590-A 51 09-MAY-1996;					
COMMENT	INNOGENETICS NV (BE)					
FEATURES	Other publication AU 3844095 960523.					
source	Location/Qualifiers					
	1..447					
	/organism="unidentified"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:32644"					
ORIGIN						
Alignment Scores:						
Pred. No.:	0.000231					Length: 447
Score:	11.00					Matches: 11
Percent Similarity:	100.00%					Conservative: 0


```

/product="core-envelope 1 protein"
/protein_id="AAU85232.1"
/db_xref="GI:52631614"
/translation="NLPGCSFISFLALLSCLTTPASAVHYKNISGIYHLTNDPNS
IYVEAGDIILHLPQVPCVTTGNTSQCVVPSPVLAVKDVTSRPGFTHVDLMVGA
AFCSALYIGDLGGVFLYQVLFTRPVHQTTQDCNCSIYTGHTGHRMAMDMN"

ORIGIN
Alignment Scores:
Pred. No.: 0.000244 Length: 474
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-851-138C-155 (1-11) x AV739423 (1-474)
QY 1 ValTYrGluAlaGlyAspIleIleLeuHisLeu 11
DB 136 GTTATGAGCGGCGTATCATCTTACACCTT 168

RESULT 5
HPCJK030A3 HPCJK030A3 1584 bp RNA linear VRL 10-FEB-1999
LOCUS Hepatitis C virus isolate JK030 gene for core, env, and part of
DEFINITION E2/NS1, partial cds.
ACCESSION D49747
VERSION D49747.1 GI:1197102
KEYWORDS core, env, and part of E2/NS1.
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (sites)
AUTHORS Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tsuda,F.,
Lesmana,L.A., Miyakawa,Y. and Mayumi,M.
TITLE Hepatitis C virus variants from Jakarta, Indonesia classifiable
into novel genotypes in the second (2e and 2f), tenth (10a) and
eleventh (11a) genetic groups
J. Gen. Virol. 77 (Pt 2), 293-301 (1996)
JOURNAL 96226020
MEDLINE 8627233
PUBMED 8627233
REFERENCE 2 (bases 1 to 1584)
AUTHORS Okamoto,H.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1584)
AUTHORS Okamoto,H.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1995) Hiroaki Okamoto, Jichi Medical School,
Immunology Division; Minamikawachi-machi, Kawachi-gun, Tochigi
329-04, Japan (E-mail:hokamoto@jichi.ac.jp,
Tel:0285-44-2111(ex.3334), Fax:0285-44-1557)
FEATURES
source
Location/Qualifiers
1..1584
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="JK030"
/db_xref="taxon:11103"
<1..339
/citation=[2]
/evidence=not_experimental
340..>1584
/codon_start=1
/product="core, env, and part of E2/NS1"
/protein_id="BAA08581.1"
/db_xref="GI:1197103"
/translation="NSTLPKQRIKNTNRPPQDKPFGGQIVGGVYVLPKPGK
GVRVVKTSRSPRSRQPIPRARTGRSWAQGYWPPLYGNGCGWAGWLLSPRG
SRPSGPNDRRRSRNLGKVIDTLTCGFADLMGYIPLVGPVGGVARALAHGVRALD
GINFATGNLPGCSFISFLALLSCLLTPAAGLEYRNASGLYVMYVMTNDCSNSIYVEAGD
IYVEAGDIILHLPQVPCVTTGNTSQCVVPSPVLAVKDVTSRPGFTHVDLMVGA
AFCSALYIGDLGGVFLYQVLFTRPVHQTTQDCNCSIYTGHTGHRMAMDMN"

ORIGIN
Alignment Scores:
Pred. No.: 0.000753 Length: 1584
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-851-138C-155 (1-11) x HPCJK030A3 (1-1584)
QY 1 ValTYrGluAlaGlyAspIleIleLeuHisLeu 11
DB 976 GTGATGAGCGCGGATATATCTCTCCATTTA 1008

RESULT 6
HPVJK070A8 HPVJK070A8 1584 bp RNA linear VRL 10-FEB-1999
LOCUS Hepatitis C virus isolate JK070 gene for core, env, and part of
DEFINITION E2/NS1, partial cds.
ACCESSION D49752
VERSION D49752.1 GI:1197162
KEYWORDS core, env, and part of E2/NS1.
SOURCE Hepatitis C virus
ORGANISM Hepatitis C virus
VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (sites)
AUTHORS Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tsuda,F.,
Lesmana,L.A., Miyakawa,Y. and Mayumi,M.
TITLE Hepatitis C virus variants from Jakarta, Indonesia classifiable
into novel genotypes in the second (2e and 2f), tenth (10a) and
eleventh (11a) genetic groups
J. Gen. Virol. 77 (Pt 2), 293-301 (1996)
JOURNAL 96226020
MEDLINE 8627233
PUBMED 8627233
REFERENCE 2 (bases 1 to 1584)
AUTHORS Okamoto,H.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1584)
AUTHORS Okamoto,H.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1995) Hiroaki Okamoto, Jichi Medical School,
Immunology Division; Minamikawachi-machi, Kawachi-gun, Tochigi
329-04, Japan (E-mail:hokamoto@jichi.ac.jp,
Tel:0285-44-2111(ex.3334), Fax:0285-44-1557)
FEATURES
source
Location/Qualifiers
1..1584
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="JK070"
/db_xref="taxon:11103"
<1..339
/citation=[2]
/evidence=not_experimental
340..>1584
/codon_start=1
/product="core, env, and part of E2/NS1"
/protein_id="BAA08586.1"
/db_xref="GI:1197163"
/translation="MSTLPKQRIKNTNRPPQDKPFGGQIVGGVYVLPKPGK
GVRVVKTSRSPRSRQPIPRARTGRSWAQGYWPPLYGNGCGWAGWLLSPRG
SRPSGPNDRRRSRNLGKVIDTLTCGFADLMGYIPLVGPVGGVARALAHGVRALD
GINFATGNLPGCSFISFLALLSCLLTPAAGLEYRNASGLYVMYVMTNDCSNSIYVEAGD
IYVEAGDIILHLPQVPCVTTGNTSQCVVPSPVLAVKDVTSRPGFTHVDLMVGA
AFCSALYIGDLGGVFLYQVLFTRPVHQTTQDCNCSIYTGHTGHRMAMDMN"

```

SOVRLPOTIFDLVAGAHGVMAGVAYYSMQGNWAKVFLVLCFLSGVDASTTISGSGA
ARSVMGITSLSFSPGSONQLVN"

ORIGIN

Alignment Scores:
Pred. No.: 0.000753 Length: 1584
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-851-138C-155 (1-11) x HPVJK070A8 (1-1584)

Qy 1 ValTYrGluAlaGlyAspIleIleLeuHisLeu 11

Db 976 GTGTATGAGCGCGGATATTATCTCCACTTG 1008

RESULT 7

AC093263/c
LOCUS Homo sapiens chromosome 5 clone RP11-259G23, complete sequence. PRI 18-DEC-2001
AC093263
AC093263
AC093263.3 GI:17921223
HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens

REFERENCE

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 141167)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 141167)
DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 141167)
DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
TITLE Submitted (18-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 18, 2001 this sequence version replaced gi:15290463.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.

FEATURES

source
1. .141167
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-259G23"

ORIGIN

Alignment Scores:
Pred. No.: 213 Length: 141167
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 72.73% Indels: 0
DB: 9 Gaps: 0

US-09-851-138C-155 (1-11) x AC093263 (1-141167)

Qy 4 AlaGlyAspIleIleLeuHisLeu 11

Db 47930 GCTGGTGATATATCTTACATCTA 47907

RESULT 8

AC016107/c
LOCUS Homo sapiens clone RP11-26L18, WORKING DRAFT SEQUENCE, 13 unordered
DEFINITION pieces.

AC016107
AC016107.3 GI:7249085
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152290)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens, clone RP11-26L18

Unpublished

2 (bases 1 to 152290)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,
Ferriere, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission

TITLE

JOURNAL Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 16, 2000 this sequence version replaced gi:6649338.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4655

Center clone name: 26.L.18

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 142723 bases at least Q40

Consensus quality: 148078 bases at least Q30

Consensus quality: 150073 bases at least Q20

Insert size: 154000; agarose-ff

Insert size: 151090; sum-of-contigs

Quality coverage: 4.3 in Q20 bases; agarose-ff

Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 235: contig of 235 bp in length
* 236 335: gap of 100 bp
* 336 2317: contig of 1982 bp in length
* 2318 2417: gap of 100 bp
* 2418 5016: contig of 2599 bp in length
* 5017 5116: gap of 100 bp
* 5117 7370: contig of 2254 bp in length

* 7371 7470: gap of 100 bp
 * 7471 15469: contig of 7999 bp in length
 * 15470 15569: gap of 100 bp
 * 15570 24842: contig of 9273 bp in length
 * 24843 24942: gap of 100 bp
 * 24943 31916: contig of 6974 bp in length
 * 31917 32016: gap of 100 bp
 * 32017 43343: contig of 11327 bp in length
 * 43344 43443: gap of 100 bp
 * 43444 52631: contig of 9188 bp in length
 * 52632 52731: gap of 100 bp
 * 52732 68745: contig of 16014 bp in length
 * 68746 84202: gap of 100 bp
 * 84203 84302: contig of 15357 bp in length
 * 84303 111032: contig of 26730 bp in length
 * 111033 111132: gap of 100 bp
 * 111133 152290: contig of 41158 bp in length.

FEATURES

source

1..152290
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="RP11-26L18"
 /clone_lib="RPCI-11 Human Male BAC"

misc_feature

1..235
 /note="assembly_fragment"
 clone end:17
 vector side:left"

misc_feature

336..2317
 /note="assembly_fragment"

misc_feature

2418..5016
 /note="assembly_fragment"

misc_feature

5117..7370
 /note="assembly_fragment"

misc_feature

7471..15469
 /note="assembly_fragment"

misc_feature

15570..24842
 /note="assembly_fragment"

misc_feature

24943..31916
 /note="assembly_fragment"

misc_feature

32017..43343
 /note="assembly_fragment"

misc_feature

43444..52631
 /note="assembly_fragment"

misc_feature

clone end:566
 vector side:left"

misc_feature

52732..68745
 /note="assembly_fragment"

misc_feature

68846..84202
 /note="assembly_fragment"

misc_feature

84303..111032
 /note="assembly_fragment"

misc_feature

111133..152290
 /note="assembly_fragment"

ORIGIN

Alignment Scores:
 Pred. No.: 229 Length: 152290
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 72.73% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-155 (1-11) x AC016107 (1-152290)

Qy 4 AlaGlyAspIleIleLeuHisLeu 11

Db 63027 GCTGTGATATAATCTTACATCTA 63004

RESULT 9

AC139271/c

ORIGIN

Alignment Scores:

Pred. No.:

238

Length:

158618

AC139271 158618 bp DNA linear HTG 29-JAN-2003
 DEFINITION Homo sapiens chromosome 5 clone RP11-586K22, WORKING DRAFT
 SEQUENCE, 9 unordered pieces.

AC139271

AC139271.1 GI:28009558

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

REFERENCE 1 (bases 1 to 158618)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS DOE Joint Genome Institute.

TITLE Sequencing of Human Chromosome 5

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 158618)

AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (29-JAN-2003) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: http://www.jgi.doe.gov

Project Information

Center Project Name: 1519790

Center clone name: RPCI-11_586K22

Summary Statistics

Consensus quality: 149962 bases at least Q40
 Consensus quality: 154153 bases at least Q30
 Consensus quality: 155835 bases at least Q20
 Estimated insert size: 175000; agarose-fp estimation
 Estimated insert size: 157818; sum-of-contigs estimation
 Quality coverage: 6.78 in Q20 bases; agarose-fp estimation
 Quality coverage: 7.52 in Q20 bases; sum-of-contigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

1 1082: contig of 1082 bp in length
 * 1083 1182: gap of unknown length
 * 1183 3928: contig of 2746 bp in length
 * 3929 4028: gap of unknown length
 * 4029 15553: contig of 11525 bp in length
 * 15554 15653: gap of unknown length
 * 15654 25408: contig of 9755 bp in length
 * 25409 25508: gap of unknown length
 * 25509 41291: contig of 15783 bp in length
 * 41292 41392: gap of unknown length
 * 41392 64482: contig of 23091 bp in length
 * 64483 89121: contig of 24539 bp in length
 * 89122 89221: gap of unknown length
 * 89222 120418: contig of 31197 bp in length
 * 120419 120518: gap of unknown length
 * 120519 158618: contig of 38100 bp in length.

FEATURES

source

1..158618
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="RP11-586K22"
 /clone_lib="RPCI human BAC library 11"

```

Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 72.73% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-155 (1-11) x AC139271 (1-158618)

Qy 4 AlaGlyAspIleLeuHisLeu 11
Db 107697 GCTGGTGATATATCTTACATCTA 107674

RESULT 10
AC139812/c 179668 bp DNA linear HTG 07-MAR-2003
LOCUS Homo sapiens chromosome 5 clone RP11-1375M22, WORKING DRAFT
DEFINITION
AC139812
VERSION AC139812.2 GI:28875956
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179668)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (13-FEB-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 179668)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Mar 7, 2003 this sequence version replaced gi:28372639.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 2747944
Center clone name: RPCI-11_1375M22
-----
Summary Statistics
Consensus quality: 179483 bases at least Q40
Consensus quality: 179537 bases at least Q30
Consensus quality: 179574 bases at least Q20
Estimated insert size: 175000; agarose-fp estimation
Estimated insert size: 179668; sum-of-contigs estimation
Quality coverage: 8.42 in Q20 bases; agarose-fp estimation
Quality coverage: 8.2 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 179668: contig of 179668 bp in length.
Location/Qualifiers
1. 179668
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-1375M22"
/clone_lib="RPCI human BAC library 11"

FEATURES
source
ORIGIN
Alignment Scores:
Pred. No.: 267 Length: 179668
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 72.73% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-155 (1-11) x AC139812 (1-179668)

Qy 4 AlaGlyAspIleLeuHisLeu 11
Db 54694 GCTGGTGATATATCTTACATCTA 54671

RESULT 11
AC119064/c 188938 bp DNA linear HTG 13-MAR-2003
LOCUS Sus scrofa clone RP44-222G1, WORKING DRAFT SEQUENCE, 10 ordered
DEFINITION pieces.
AC119064
VERSION AC119064.4 GI:289333583
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 188938)
AUTHORS Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
Carliaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
Haghghi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
Pagirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
Reddix-Dugue,N., Schandler,K., Schueler,W.G., Sison,C.,
Stantripoop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
REFERENCE 2 (bases 1 to 188938)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (24-APR-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE 3 (bases 1 to 188938)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
On Mar 13, 2003 this sequence version replaced gi:27884877.
-----Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
-----Project Information
Center project name: dbc
Center clone name: 222G01

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8x average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.

```



```

Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 72.73% Indels: 0
DB: 2 Gaps: 0

US-09-851-138c-155 (1-11) x AC139804 (1-189043)

Qy 4 AlaGlyAspIleLeuHisLeu 11
|||||
Db 11739 GCTGGTGATATATCTTACATCTA 11716

RESULT 13
AC139462/c
LOCUS AC139462 196706 bp DNA linear HTG 04-FEB-2003
DEFINITION Homo sapiens chromosome 5 clone RP11-1223D8, WORKING DRAFT
AC139462
AC139462.1 GI:28201492
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 196706)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
REFERENCE 2 (bases 1 to 196706)
AUTHORS Unpublished
JOURNAL
TITLE Direct Submission
JOURNAL
SUBMITTED (04-FEB-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 2689346
Center clone name: RPCI-11_1223D8
-----
Summary Statistics
Consensus quality: 195628 bases at least Q40
Consensus quality: 195773 bases at least Q30
Consensus quality: 195935 bases at least Q20
Estimated insert size: 175000; agarose-fp estimation
Estimated insert size: 196506; sum-of-contigs estimation
Quality coverage: 7.88 in Q20 bases; agarose-fp estimation
Quality coverage: 7.02 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6223: contig of 6223 bp in length
* 6224 6323: gap of unknown length
* 6324 42059: contig of 35736 bp in length
* 42060 42159: gap of unknown length
* 42160 196706: contig of 154547 bp in length.
Location/Qualifiers
1. 196706
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-1223D8"
/clone_lib="RPCI human BAC library 11"

FEATURES
source
ORIGIN

```

```

Alignment Scores:
Pred. No.: 290 Length: 196706
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 72.73% Indels: 0
DB: 2 Gaps: 0

US-09-851-138c-155 (1-11) x AC139462 (1-196706)

Qy 4 AlaGlyAspIleLeuHisLeu 11
|||||
Db 10100 GCTGGTGATATATCTTACATCTA 10077

RESULT 14
AC139814
LOCUS AC139814 198361 bp DNA linear HTG 13-FEB-2003
DEFINITION Homo sapiens chromosome 5 clone RP11-1384I17, WORKING DRAFT
AC139814
AC139814.1 GI:28372641
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 198361)
DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
REFERENCE 2 (bases 1 to 198361)
AUTHORS Unpublished
JOURNAL
TITLE Direct Submission
JOURNAL
SUBMITTED (13-FEB-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 2751299
Center clone name: RPCI-11_1384I17
-----
Summary Statistics
Consensus quality: 196122 bases at least Q40
Consensus quality: 196639 bases at least Q30
Consensus quality: 197164 bases at least Q20
Estimated insert size: 175000; agarose-fp estimation
Quality coverage: 4.45 in Q20 bases; agarose-fp estimation
Quality coverage: 3.94 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2195: contig of 2195 bp in length
* 2196 2295: gap of unknown length
* 2296 5330: contig of 3035 bp in length
* 5331 5430: gap of unknown length
* 5431 34677: contig of 29247 bp in length
* 34678 34777: gap of unknown length
* 34778 65756: contig of 30979 bp in length
* 65757 65856: gap of unknown length
* 65857 99292: contig of 33436 bp in length
* 99293 99392: gap of unknown length
* 99393 142332: contig of 42940 bp in length
* 142333 142432: gap of unknown length
* 142433 198361: contig of 55929 bp in length.

```

FEATURES
source

Location/Qualifiers
1. .198361
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-1384117"
/clone_lib="RPC1 human BAC library 11"

ORIGIN

Alignment Scores:
Pred. No.: 293 Length: 198361
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 72.73% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-155 (1-11) x AC139814 (1-198361)

Qy 4 AlaGlyAspIleIleLeuHisLeu 11

Db 97776 GCTGTGATATATCTTACATCTA 97799

RESULT 15

CQ482459/c
LOCUS CQ482459 446 bp DNA linear PAT 30-JAN-2004
DEFINITION Sequence 14326 from Patent WO0160860.
ACCESSION CQ482459
VERSION CQ482459.1 GI:41448078
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Schlegel, R., Endege, W.O. and Monahan, J.E.
TITLE Genes differentially expressed in human prostate cancer and their use
JOURNAL Patent: WO 0160860-A 14326 23-AUG-2001;
Millennium Predictive Medicine, Inc. (US)

FEATURES
source

Location/Qualifiers
1. .446
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 16 Length: 446
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 63.64% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-155 (1-11) x CQ482459 (1-446)

Qy 1 ValTyrGluAlaGlyAspIle 7

Db 437 GTCTATGAAGCAGGGGATATT 417

Search completed: March 3, 2005, 18:34:52
Job time : 534.072 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 14:30:42 ; Search time 85.6 Seconds
(without alignments)
829.870 Million cell updates/sec

Title: US-09-851-138c-138
Perfect score: 12
Sequence: 1 LEYNASGLYMW 12

Scoring table: OLIGO*
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8769587

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlp
-O=/cgn2_1/USPTO spool_p/US09851138/runat_28022005_120306_21457/app_query.fasta_1.1123
-DB=N_Geneseq_16Dec04 -QNT=fastap -SUFFIX=olg.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09851138 @CGN 1 1.1418 @runat_28022005_120306_21457 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12	100.0	447	2	Aat27962 Hepatitis
c 2	8	66.7	9444	2	Aat13279 CDNA to g
3	7	58.3	298	6	Ab174296 Corn tabs
c 4	7	58.3	321	11	Abd03324 Pseudomon
5	7	58.3	577	2	Aaq35077 HCV envel

ACA36263	Prokaryot
ACH95253	Klebsiell
ACH95218	Klebsiell
ABL15605	Drosophil
AAQ63499	Blood tra
ABL15604	Drosophil
Continuation (5 of	
ACN44806	Human gen
AZ02178	PCR prime
AAA05618	PCR prime
ACD79681	E. coli K
ACD74487	E. coli K
ACD80740	E. coli K
AAA69387	Bacteriop
ADL87537	DNA up-re
ADL87538	DNA up-re
AAC21197	Human sec
AAA69233	Bacteriop
AZA14444	Human gen
ABV35765	Human pro
ABV00094	Human pro
ABL37519	Human col
ACH30848	Human bon
AAV21207	Mouse PPA
AAK56814	Human imm
ABX51473	Bovine ES
ABV98650	Human pan
ADL87662	DNA up-re
ADL87661	DNA up-re
ABN26365	Human ORF
AAH28963	Drosophil
ACA32012	Prokaryot
ABV14687	Human pro
ADL13352	Human ste
AAI28994	Colon tum
ABZ33180	Human col
AAV52657	Human nat
ABV30439	Human pro
ABV39413	Human pro
ABL93430	Arabidops

ALIGNMENTS

RESULT 1
AAT27962
ID AAT27962 standard; DNA; 447 BP.
XX
AC AAT27962;
XX
DT 11-MAR-1997 (first entry)
XX
DE Hepatitis C virus type 10a isolate NN98 bases 478-925.
XX
KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
PCR; primer; probe; antibody; infection; ss.
XX
OS Hepatitis C virus.
XX
PN WO9613590-A2.
XX
PD 09-MAY-1996.
XX
PF 23-OCT-1995; 95WO-EP004155.
XX
PR 21-OCT-1994; 94EP-00870166.
PR 28-JUN-1995; 95EP-00870076.
XX
(INNO-) INNOGENETICS NV.
XX
PI Maertens G, Stuyver L;
XX
DR WPI; 1996-251460/25.

DR P-PSDB; AAR96551.
 XX Hepatitis C virus poly:nucleic acid unique to unidentified sub:type -
 PT used to develop probes and primers for new sub:types and vaccines to
 PT prevent and treat infection.
 XX
 PS Claim 6; Fig 3; 150pp; English.
 XX The sequences AAT27937-T27989 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f, 4a-
 CC j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l, 3g,
 CC 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
 CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the genome.
 CC This sequence represents nucleotides 478-925 from the HCV type 10a
 CC isolate NE98. The new HCV types were isolated from patients with chronic
 CC HCV from the Benelux countries, France, Cameroon and Vietnam, because of
 CC their aberrant reactivities. The RNA was extracted, cDNA synthesised and
 CC PCR amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions
 CC were sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences. The
 CC sequences were used to generate the peptides AAR96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect anti
 CC -HCV antibodies, for HCV typing or to prevent HCV infections
 XX
 SQ Sequence 447 BP; 82 A; 130 C; 114 G; 118 T; 0 U; 3 Other;
 Alignment Scores:
 Pred. No.: 0.000152 Length: 447
 Score: 12.00 Matches: 12
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-138 (1-12) x AAT27962 (1-447)
 QY 1 LeuGluTyrArgAsnAlaSerGlyLeuTyrMetVal 12
 DB 97 CTGGAGTACCGTAATGCTCCGGACTCTACATGGTA 132

RESULT 2
 AAT13279/c
 ID AAT13279 standard; cDNA; 9444 BP.
 XX
 AC AAT13279;
 XX
 DT 20-SEP-1996 (first entry)
 XX
 DE cDNA to genomic hepatitis C virus RNA.
 XX
 KW hepatitis C virus; antibody; detection; diagnosis; vaccine; classify;
 KW subtype; ss.
 XX
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT CDS complement(34..9105)
 FT /*tag= a
 XX
 FN JF08056672-A.
 XX
 XX
 PD 05-MAR-1996.
 XX
 XX
 PF 26-AUG-1994; 94JP-00223933.
 XX
 XX 26-AUG-1994; 94JP-00223933.
 PR
 XX (SAYA/) SAYAMA K.
 PA
 XX WPI; 1996-182301/19.
 DR P-PSDB; AAR94462.
 XX

PT Hepatitis C virus genomic RNA, DNA and related proteins - useful for
 PT detection, diagnosis and identification of hepatitis C virus sub-type.
 XX
 PS Claim 2; Page 9-12; 25pp; Japanese.
 XX
 CC The present sequence represents cDNA to a hepatitis C virus (HCV) genomic
 CC RNA. The sequence encodes a polypeptide contg. a 3023 amino acid sequence
 CC (see AAR94462) which can be easily detected by antibodies in an assay for
 CC the detection of HCV. The DNA and the protein are useful for classifying
 CC the subtype of HCV. At least a part of the protein may be used as a
 CC vaccine against HCV
 XX
 SQ Sequence 9444 BP; 2079 A; 2608 C; 2682 G; 2075 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 68.1 Length: 9444
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 66.67% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-138 (1-12) x AAT13279 (1-9444)
 QY 1 LeuGluTyrArgAsnAlaSerGly 8
 DB 8532 CTGGAGTACAGGAATCGTCTGCG 8509

RESULT 3
 ABL74296
 ID ABL74296 standard; cDNA; 298 BP.
 XX
 AC ABL74296;
 XX
 DT 14-MAY-2002 (first entry)
 XX
 DE Corn tassel-derived polynucleotide (cdps) SEQ ID NO:3670.
 XX
 KW Corn; corn tassel-derived polynucleotide; cdps; hybrid breeding; CDPs;
 KW inheritance; characteristic; growth; development; disease resistance;
 KW environmental adaptability; quality; yield; molecular marker;
 KW multigene trait; plant breeding; corn tassel; gene; ss.
 XX
 OS Zea mays.
 XX
 PN US2001051335-A1.
 PD 13-DEC-2001.
 XX
 PF 16-APR-1999; 99US-00294093.
 XX
 PR 21-APR-1998; 98US-0082567P.
 XX
 PA (LALG/) LALGUDI R V.
 PA (ITOL/) ITO L Y.
 PA (SHER/) SHERMAN B K.
 XX
 PI Lalgudi RV, Ito LY, Sherman BK;
 XX
 DR WPI; 2002-163647/21.
 XX
 XX Novel purified corn tassel-derived polynucleotide useful for determining
 PT altered gene expression, to recover regulatory elements and to follow
 PT inheritance of desirable characteristics through hybrid breeding
 PT programs.
 XX
 PS Claim 1; SEQ ID NO 3670; 201pp; English.
 XX
 CC The present sequence describes a purified corn tassel-derived
 CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
 CC selected from those given in ABL70627 to ABL76833. The cdps sequences
 CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)
 CC can be used for determining altered gene expression, to recover

CC regulatory elements and to follow inheritance of desirable
 CC characteristics through hybrid breeding programs. (I) are also useful in
 CC the evaluation, and alteration of desired characteristics associated with
 CC growth and development, disease resistance, environmental adaptability,
 CC quality and yield, and as molecular markers for studying inheritance of
 CC multigene traits in a plant breeding program. (I) can be used to produce
 CC a tassel-specific profile of gene transcription, a transcript image, to
 CC clone regulatory elements for use in transformation vectors, to express a
 CC polypeptide, to identify, isolate or extend identical or related corn
 CC tassel nucleic acid sequences from DNA libraries, in nucleic acid
 CC hybridisation or amplification technologies, as query sequences to
 CC determine homology of known sequences, as probe for use in Southern or
 CC Northern hybridisation, and to identify the presence of and/or to
 CC determine the degree of similarity between two (or more) nucleic acid
 CC sequences

SQ Sequence 298 BP; 60 A; 79 C; 78 G; 74 T; 0 U; 7 Other;

Alignment Scores: 35.4 Length: 298
 Pred. No.: 7.00 Matches: 7
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 100.00% Indels: 0
 Query Match: 58.33% Gaps: 0
 DB: 6

US-09-851-138C-138 (1-12) x ABL74296 (1-298)

Qy 3 TyrArgAsnAlaSerGlyLeu 9
 |||||
 Db 163 TATCGCAATGCTCTGGGTTG 183

RESULT 4
 ABD03324/c
 ID ABD03324 standard; DNA; 321 BP.

AC ABD03324;
 XX 29-JUL-2004 (first entry)
 DT Pseudomonas aeruginosa polynucleotide #1928.

DE Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
 KW antibacterial.

XX Pseudomonas aeruginosa.

OS US5551795-B1.

PN 22-APR-2003.

PD 18-FEB-1999; 99US-00252991.

PF 18-FEB-1998; 98US-0074788P.

PR 27-JUL-1998; 98US-0094190P.

XX (GENO-) GENOME THERAPEUTICS CORP.

PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;

XX WPI; 2003-615309/58.

DR P-PsDB; ABO69753.

XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of
 PT pathological conditions resulting from bacterial infection.

XX Disclosure; SEQ ID NO 1928; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the
 CC polynucleotides encoding them. The sequences are useful in diagnosis and
 CC therapy of pathological conditions, as molecular targets for diagnostics,
 CC prophylaxis and treatment of pathological conditions resulting from a

CC bacterial infection, for evaluating a compound, such as a polypeptide,
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of
 CC effective antibacterial targets, as targets for antibacterial drugs,
 CC including anti-P. aeruginosa drugs, as templates for recombinant
 CC production of P. aeruginosa-derived peptides or polypeptides, as target
 CC components for diagnosis and/or treatment of P. aeruginosa-caused
 CC infection, and in detection of P. aeruginosa sequences or other sequences
 CC of Pseudomonas species using biochip technology. Sequences ABD01397-
 CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format from USPTO at
 CC seqdata.uspto.gov/sequence.html

XX SQ Sequence 321 BP; 78 A; 93 C; 92 G; 58 T; 0 U; 0 Other;

Alignment Scores: 37.9 Length: 321
 Pred. No.: 7.00 Matches: 7
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 100.00% Indels: 0
 Query Match: 58.33% Gaps: 0
 DB: 11

US-09-851-138C-138 (1-12) x ABD03324 (1-321)

Qy 4 ArgAsnAlaSerGlyLeuTyr 10
 |||||
 Db 115 CGAAACGCATCAGGGTTATAT 95

RESULT 5
 AAQ35077
 ID AAQ35077 standard; DNA; 577 BP.

AC AAQ35077;
 XX 20-MAY-1993 (first entry)
 DT HCV envelope region probe 1.

DE Envelope; region; type C; hepatitis; virus; HCV; vaccine; serum;
 KW non-A, non-B; ss.

XX Synthetic.

OS JP04349885-A.

PN 04-DEC-1992.

PD 29-MAY-1991; 91JP-00152169.

PF 29-MAY-1991; 91JP-00152169.

PR (TEIJ) TEIJIN LTD.

XX WPI; 1993-022708/03.

DR Envelope region nucleic acid fragment - for type C hepatitis virus (I),
 PT for producing vaccine.

XX Disclosure; Page 9; 13pp; Japanese.

XX The sequences given in AAQ35077-89 are probes which were used to in the
 CC isolation of a novel nucleic acid encoding an envelope region of type C
 CC hepatitis virus (HCV). The isolated fragment can be used for the
 CC preparation of a vaccine for hepatitis C. The envelope region DNA for was
 CC derived from the serum of non-A, non-B hepatitis patients

XX SQ Sequence 577 BP; 97 A; 179 C; 162 G; 139 T; 0 U; 0 Other;

Alignment Scores: 65.2 Length: 577
 Pred. No.: 7.00 Matches: 7
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00%

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-138 (1-12) x AAQ35077 (1-577)

QY 4 ArgAsnAlaSerGlyLeuTyr 10

DB 139 CGCAACGCGTCCGGGTGTAC 159

RESULT 6

ACA36263
ID ACA36263 standard; DNA; 2286 BP.

XX AC ACA36263;

DT 19-JUN-2003 (first entry)

XX DE Prokaryotic essential gene #17920.

XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX OS Klebsiella pneumoniae.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362659P.

XX PA (BLIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR P-PSDB; ABU32393.

XX PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 14; SEQ ID NO 24133; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 2286 BP; 520 A; 704 C; 617 G; 445 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 234 Length: 2286
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 8 Gaps: 0

US-09-851-138C-138 (1-12) x ACA36263 (1-2286)

QY 4 ArgAsnAlaSerGlyLeuTyr 10

DB 561 CGAAACGCGTCTGGAATATAT 581

RESULT 7

ACH95253/c

ID ACH95253 standard; DNA; 2400 BP.

XX AC ACH95253;

DT 29-JUL-2004 (first entry)

DE Klebsiella pneumoniae polynucleotide seqid 1048.

XX KW Recombinant expression vector; transcription regulatory element;
KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.
XX OS Klebsiella pneumoniae.
XX PN US6610836-B1.
XX PD 26-AUG-2003.
XX PF 27-JAN-2000; 2000US-00489039.
XX PR 29-JAN-1999; 99US-0117747P.

XX PA (GENO-) GENOME THERAPEUTICS CORP.

XX PI Breton GL, Osborne M;

XX WPI; 2003-895346/82.

XX P-PSDB; ABO61702.

XX PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
XX preparing a vaccine composition against Klebsiella pneumoniae.
XX PS Disclosure; SEQ ID NO 1048; 932pp; English.

XX CC The invention describes a new isolated nucleic acid encoding a Klebsiella
CC pneumoniae polypeptide. Also described are: a recombinant expression
CC vector comprising the nucleic acid, operably linked to a transcription
CC regulatory element; and a cell comprising the recombinant expression
CC vector. The nucleic acid is useful for preparing a vaccine composition
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella
CC pneumoniae polypeptide of the invention

SQ Sequence 2400 BP; 471 A; 643 C; 739 G; 547 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 244 Length: 2400

Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 11 Gaps: 0

US-09-851-138C-138 (1-12) x ACH95253 (1-2400)

Qy 4 ArgAnAlaSerGlyLeuTyr 10
|||
Db 1822 CGAAACGGCTCTGGACTATAT 1802

RESULT 8
ACH95218
ID ACH95218 standard; DNA; 2454 BP.
XX
AC ACH95218;
XX
DT 29-JUN-2004 (first entry)
XX
DE Klebsiella pneumoniae polynucleotide seqid 1013.
XX
KW Recombinant expression vector; transcription regulatory element;
KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.
XX
OS Klebsiella pneumoniae.
XX
PN US6610836-B1.
XX
PD 26-AUG-2003.
XX
PF 27-JAN-2000; 2000US-00489039.
XX
PR 29-JAN-1999; 99US-0117747P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton GL, Osborne M;
XX
DR WPI; 2003-8953346/82.
DR P-PSDB; ABO61667.
XX
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS Disclosure; SEQ ID NO 1013; 932pp; English.
XX
CC The invention describes a new isolated nucleic acid encoding a Klebsiella
CC pneumoniae polypeptide. Also described are: a recombinant expression
CC vector comprising the nucleic acid, operably linked to a transcription
CC regulatory element; and a cell comprising the recombinant expression
CC vector. The nucleic acid is useful for preparing a vaccine composition
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella
CC pneumoniae polypeptide of the invention
XX
SQ Sequence 2454 BP; 574 A; 746 C; 639 G; 495 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 250 Length: 2454
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 11 Gaps: 0

US-09-851-138C-138 (1-12) x ACH95218 (1-2454)

Qy 4 ArgAnAlaSerGlyLeuTyr 10
|||
Db 726 CGAAACGGCTCTGGACTATAT 746

RESULT 9
ABL15605

ID ABL15605 standard; cDNA; 5279 BP.
XX
AC ABL15605;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 41297.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB71502.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 41297; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 5279 BP; 1391 A; 1425 C; 1339 G; 1124 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 508 Length: 5279
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-138 (1-12) x ABL15605 (1-5279)

Qy 4 ArgAnAlaSerGlyLeuTyr 10
|||
Db 3783 CGTAACGCGACGGGATTATAC 3803

RESULT 10
AAQ63499
ID AAQ63499 standard; cDNA; 9436 BP.
XX
AC AAQ63499;
XX
DT 17-JAN-1995 (first entry)
XX
DE Blood transmissible NANBHV genome.
XX
KW Polymerase chain reaction; PCR; amplify; primer; non-A, non-B hepatitis;

KW NANBH; virus; blood transmissible; detection; hepatitis virus; RT-PCR;
 KW C100 antibody; HCV RNA; NS5 region; ds.
 XX OS Non-A.
 OS non-B hepatitis virus.
 XX FH Key Location/Qualifiers
 FT CDS 342..9374
 FT /*tag= a
 XX
 PN JP06105690-A.
 XX
 XX 19-APR-1994.
 PD
 XX 10-MAR-1992; 92JP-00051885.
 PF
 XX 10-MAR-1992; 92JP-00051885.
 PR
 XX (KAEN/) KAENNO K.
 PA
 XX WPI; 1994-163130/20.
 XX DR P-PSDB; AAR53417.
 DR
 XX
 XX Blood-transmissible non-A non-B hepatitis virus DNA - used for detection
 PT of hepatitis virus.
 PT
 XX Claim 1; Page 8-20; 22pp; Japanese.
 PS
 XX This sequence represents the genome of a blood transmissible non-A, non-B
 CC hepatitis (NANBH) virus. This sequence was isolated using the primers
 CC given in AAQ63500-35. The amplified fragments are used in the detection
 CC of hepatitis virus. This target DNA was isolated from serum of
 CC chronically infected NANBH patients who were C100 antibody-positive and
 CC HCV RNA (NS5 region) positive. Reverse transcription-PCR and PCR were
 CC performed on cDNA and the total human NANBH DNA was constructed from 23
 CC clones
 CC
 XX Sequence 9436 BP; 1876 A; 2840 C; 2655 G; 1974 T; 0 U; 91 Other;
 SQ
 Alignment Scores:
 Pred. No.: 870 Length: 9436
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 58.33% Indels: 0
 DB: 2 Gaps: 0
 US-09-851-138C-138 (1-12) x AAQ63499 (1-9436)
 QY 4 ArgAsnAlaSerGlyLeuTyr 10
 Db 924 CGCAACGCATCCGGCTGTAC 944
 RESULT 11
 ABL15604/c
 ID ABL15604 standard; cDNA; 20811 BP.
 XX
 AC ABL15604;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 41294.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 KW
 XX Drosophila melanogaster.
 OS
 PN WO200171042-A2.
 XX
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US009231.
 PF

XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI; 2001-656860/75.
 DR P-PSDB; ABB71501.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 PT
 XX Claim 1; SEQ ID NO 41294; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABLI6176-ABL30511), expressed DNA
 CC sequences (ABLI01840-ABLI6175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 20811 BP; 5661 A; 4171 C; 4817 G; 6162 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.81e+03 Length: 20811
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 58.33% Indels: 0
 DB: 4 Gaps: 0
 US-09-851-138C-138 (1-12) x ABLI5604 (1-20811)
 QY 4 ArgAsnAlaSerGlyLeuTyr 10
 Db 6455 CGTAACGCAAGCGATTATAC 6435
 RESULT 12
 ABD32594_4
 Continuation (5 of 7) of ABD32594 from base 400001 (Mouse cancer-associated genomic DNA N
 WP Sequence split into 7 fragments LOCUS ABD32594 Accession ABD32594
 WP Fragment Name Begin End
 WP ABD32594_0 1 110000
 WP ABD32594_1 100001 210000
 WP ABD32594_2 200001 310000
 WP ABD32594_3 300001 410000
 WP ABD32594_4 400001 510000
 WP ABD32594_5 500001 610000
 WP ABD32594_6 600001 684187
 Alignment Scores:
 Pred. No.: 8.47e+03 Length: 110000
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 58.33% Indels: 0
 DB: 13 Gaps: 0
 US-09-851-138C-138 (1-12) x ABD32594_4 (1-110000)
 QY 3 TyrArgAsnAlaSerGlyLeu 9
 Db 81798 TATAGGAATGCTTCAGGCCTT 81818
 RESULT 13
 ACN44806

ID ACN44806 standard; DNA; 177587 BP.
 AC ACN44806;
 XX
 DT 18-NOV-2004 (first entry)
 DE Human genomic sequence HCG40093.
 XX
 KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2003073826-A2.
 XX
 PD 12-SEP-2003.
 XX
 PF 28-FEB-2003; 2003WO-US006235.
 XX
 PR 01-MAR-2002; 2002US-00087192.
 XX
 PA (SAGR-) SAGRES DISCOVERY.
 XX
 PI Morris DW;
 XX
 DR WPI; 2003-328604/31.
 XX
 PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
 PT comprises a nucleotide sequence.
 PS Claim 1; SEQ ID NO 1438; Opp; English.
 XX
 CC The present invention relates to novel DNA and protein sequences which
 CC are associated with carcinomas. The sequences are useful for: (i) for
 CC screening drug candidates; (ii) for screening of bioactive agent capable
 CC of binding to Carcino Associated Protein (CAP); (iii) for screening of
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
 CC determining Carcino Associated (CA) gene copy number. In addition, the
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
 CC carcinoma including lymphoma. The present sequence is one such CA coding
 CC sequence. Note: This patent is an equivalent to basic patent
 CC US2002182586A1, for which no sequence data was published
 XX
 SQ Sequence 177587 BP; 49045 A; 38259 C; 39386 G; 50877 T; 0 U; 20 Other;
 Alignment Scores:
 Pred. No.: 1.32e+04 Length: 177587
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 58.33% Indels: 0
 DB: 11 Gaps: 0
 US-09-851-138c-138 (1-12) x ACN44806 (1-177587)
 QY 5 AsnAlaSerGlyLeuTyrMet 11
 DB 30644 AATGCTTCGGGTATACATG 30664
 RESULT 14
 AA202178/c
 ID AA202178 standard; DNA; 20 BP.
 XX
 AC AA202178;
 XX
 DT 07-OCT-1999 (first entry)
 DE PCR primer used to amplify an ORF of Chlamydia trachomatis.
 XX
 DE Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW

KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis; PCR primer;
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
 XX
 OS Synthetic.
 OS Chlamydia trachomatis.
 XX
 PN WO9928475-A2.
 XX
 PD 10-JUN-1999.
 XX
 PF 27-NOV-1998; 98WO-IB001939.
 XX
 PR 28-NOV-1997; 97FR-00015041.
 PR 17-DEC-1997; 97FR-00016034.
 PR 04-NOV-1998; 98US-0107077P.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffais R;
 XX
 DR WPI; 1999-371125/31.
 XX
 PT Genome sequence of Chlamydia trachomatis.
 XX
 PS Disclosure; Page 1503; 1755pp; English.
 XX
 CC PCR primers AA201426-206209 were used to amplify open reading frames
 CC (ORFs) of the genome of Chlamydia trachomatis (see AA201425). These ORFs
 CC encode polypeptides (see AAY36754-Y37949) which can be used as vaccines
 CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
 CC be used to control growth of the microorganism. Chlamydia trachomatis is
 CC responsible for a large number of diseases, e.g. eye diseases such as
 CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
 CC conjunctivitis; genital diseases such as nongonococcal urethritis;
 CC epididymitis, cervicitis, salpingitis, perihepatitis, bartholinitis;
 CC pneumopathy in breast feeding infants, and venereal lymphogranulomatosis.
 CC The polypeptides of the invention may be of use in treating these
 CC diseases
 XX
 SQ Sequence 20 BP; 2 A; 7 C; 4 G; 7 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 36.9 Length: 20
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 50.00% Indels: 0
 DB: 2 Gaps: 0
 US-09-851-138c-138 (1-12) x AA202178 (1-20)
 QY 3 TyrArgAsnAlaSerGly 8
 DB 18 TACAGGAATGCGACCGA 1
 RESULT 15
 AAA05618
 ID AAA05618 standard; DNA; 35 BP.
 XX
 AC AAA05618;
 XX
 DT 05-JUN-2000 (first entry)
 DE PCR primer SAV-1, SEQ ID NO:12.
 XX
 KW Phage display; bacteriophage M13; fusion protein; major coat protein;
 KW protein VIII; phagemid vector; electroporation; combinatorial library;
 KW PCR primer; ss.
 XX
 OS Streptomyces avidinii.
 XX
 PN WO200006717-A2.

```

XX 10-FEB-2000.
XX
XX PF 22-JUL-1999; 99WO-US016596.
XX
XX PR 27-JUL-1998; 98US-0094291P.
XX 08-OCT-1998; 98US-0103514P.
XX 10-MAY-1999; 99US-0133296P.
XX 19-MAY-1999; 99US-0134870P.
XX
XX (GETH ) GENENTECH INC.
XX
XX PI Sidhu SS, Weiss GA, Wells JA;
XX
XX DR WPI; 2000-181122/16.
XX
XX Fusion proteins comprising a heterologous protein and a viral variant
XX major coat protein useful in phage display systems for improving
XX transformation efficiency.
XX
XX Example 9; Page 68; 118pp; English.
XX
XX The invention relates to novel fusion proteins comprising a heterologous
XX polypeptide fused to a variant (non-wild type) bacteriophage major coat
XX protein (protein VIII). The invention also relates to replicable
XX expression vectors which contain a gene encoding the fusion protein; host
XX cells containing the expression vectors; phages which display the fusion
XX protein on their surface; phage libraries displaying a plurality of
XX different fusion proteins on viral surfaces; and methods of using these
XX compositions. The fusion proteins the invention are well tolerated in
XX phage display systems. Variants of the major coat proteins can be used to
XX alter the number of fusion proteins incorporated into a virus particle.
XX Hyper-functional variants can be used to increase the number of fusion
XX proteins incorporated into a virus particle. Conversely, hypo-functional
XX variants can be used to decrease fusion protein incorporation. This is
XX useful for tailoring the incorporation of fusion proteins into virus
XX particles to achieve a desired level of valency. The variant replicable
XX plasmid/phagemid vectors are useful for producing polypeptides of
XX interest. The methods are useful for improving the transformation of
XX cells by highly purifying DNA. The present invention uses affinity DNA
XX purification to reduce ionic impurities and thus reduce the conductance
XX associated with a unit mass of DNA. This is an advantageous in
XX electroporation methods for increasing the concentration of DNA present.
XX The increase in DNA entering the host cell provides a greater number of
XX transformants per electroporation and allows one to prepare larger
XX combinatorial libraries which overcomes the prior art problem of small
XX library size using recombinant DNA. Sequences AAA05615-A05621 represent
XX PCR primers used in the exemplifications of the invention for phagemid
XX construction
XX
XX SQ Sequence 35 BP; 7 A; 11 C; 11 G; 6 T; 0 U; 0 Other;

```

```

Alignment Scores:
Pred. No.: 62 Length: 35
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.00% Indels: 0
DB: 3 Gaps: 0

```

US-09-851-138C-138 (1-12) x AAA05618 (1-35)

```

Qy 3 TyrArgAsnAlaSerGly 8
Db 4 TATCGGAATGATCGGCG 21

```

Search completed: March 3, 2005, 16:26:05
Job time : 105.6 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:54:32 ; Search time 24.7385 Seconds
(without alignments)
793.716 Million cell updates/sec

Title: US-09-851-138c-138
Perfect score: 12
Sequence: 1 LEYRNAGLYMV 12

Scoring table: OLIGO-3
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2396311

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO/epool_p/US09851138/runat_28022005_120307_21495/app_query.fasta_1.1123
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=oligo.rni -MINWATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MAYRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09851138 @CGN 1 1 249 @runat_28022005_120307_21495 -NCPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A COMB.seq:
2: /cgn2_6/ptodata/1/ina/5B COMB.seq:
3: /cgn2_6/ptodata/1/ina/6A COMB.seq:
4: /cgn2_6/ptodata/1/ina/6B COMB.seq:
5: /cgn2_6/ptodata/1/ina/PTUS COMB.seq:
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	100.0	447	3	US-08-836-075A-51
C 2	7	58.3	321	4	US-09-851-138c-138
C 3	7	58.3	601	4	US-09-252-991A-1328
C 4	7	58.3	601	4	US-09-949-016-63514
C 5	7	58.3	699	4	US-09-949-016-169481
C 6	7	58.3	2400	4	US-09-270-767-12425
C 7	7	58.3	2454	4	US-09-489-039A-1048
C 8	7	58.3	104077	4	US-09-489-039A-1013
C 9	7	58.3	160759	4	US-09-949-016-13593
C 10	6	50.0	186	4	US-09-949-016-16514
C 11	6	50.0	210	4	US-09-513-999C-25272
C 12	6	50.0	373	3	US-08-917-653-3

C 13	6	50.0	431	4	US-09-976-594-1081	Sequence 1081, Ap
14	6	50.0	458	3	US-08-927-219-40	Sequence 40, Appl
15	6	50.0	474	4	US-09-902-540-8014	Sequence 8014, Ap
16	6	50.0	531	4	US-09-252-991A-4374	Sequence 4374, Ap
17	6	50.0	574	4	US-09-878-281A-118	Sequence 118, App
18	6	50.0	574	4	US-09-878-281A-1122	Sequence 122, App
19	6	50.0	576	1	US-08-086-428B-40	Sequence 40, Appl
20	6	50.0	576	1	US-08-086-428B-41	Sequence 41, Appl
21	6	50.0	576	1	US-08-086-428B-42	Sequence 42, Appl
22	6	50.0	576	1	US-08-086-428B-45	Sequence 45, Appl
23	6	50.0	576	1	US-08-086-428B-46	Sequence 46, Appl
24	6	50.0	576	1	US-08-086-428B-47	Sequence 47, Appl
25	6	50.0	576	1	US-08-086-428B-48	Sequence 48, Appl
26	6	50.0	576	1	US-08-086-428B-49	Sequence 49, Appl
27	6	50.0	576	1	US-08-086-428B-50	Sequence 50, Appl
28	6	50.0	576	2	US-08-468-570-40	Sequence 40, Appl
29	6	50.0	576	2	US-08-468-570-41	Sequence 41, Appl
30	6	50.0	576	2	US-08-468-570-42	Sequence 42, Appl
31	6	50.0	576	2	US-08-468-570-45	Sequence 45, Appl
32	6	50.0	576	2	US-08-468-570-46	Sequence 46, Appl
33	6	50.0	576	2	US-08-468-570-47	Sequence 47, Appl
34	6	50.0	576	2	US-08-468-570-48	Sequence 48, Appl
35	6	50.0	576	2	US-08-468-570-49	Sequence 49, Appl
36	6	50.0	576	2	US-08-468-570-50	Sequence 50, Appl
37	6	50.0	576	2	US-08-290-665A-40	Sequence 40, Appl
38	6	50.0	576	2	US-08-290-665A-41	Sequence 41, Appl
39	6	50.0	576	2	US-08-290-665A-42	Sequence 42, Appl
40	6	50.0	576	2	US-08-290-665A-45	Sequence 45, Appl
41	6	50.0	576	2	US-08-290-665A-46	Sequence 46, Appl
42	6	50.0	576	2	US-08-290-665A-47	Sequence 47, Appl
43	6	50.0	576	2	US-08-290-665A-48	Sequence 48, Appl
44	6	50.0	576	2	US-08-290-665A-49	Sequence 49, Appl
45	6	50.0	576	2	US-08-290-665A-50	Sequence 50, Appl

ALIGNMENTS

RESULT 1
US-08-836-075A-51
; Sequence 51, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836, 075A
FILING DATE: 21 Apr 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:

```

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307

```

```
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 12425
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-12425

Alignment Scores:
Pred. No.: 13.5 Length: 699
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-138 (1-12) x US-09-270-767-12425 (1-699)

QY 4 ArgAsnAlaSerGlyLeuTyr 10
DB 33 CGTAACGCGGCGGATTATAC 13

RESULT 6
US-09-489-039A-1048/c
; Sequence 1048, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1048
; LENGTH: 2400
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1048

Alignment Scores:
Pred. No.: 45 Length: 2400
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-138 (1-12) x US-09-489-039A-1048 (1-2400)

QY 4 ArgAsnAlaSerGlyLeuTyr 10
DB 1822 CGAAACGCGCTGGACTATAT 1802

RESULT 7
US-09-489-039A-1013
; Sequence 1013, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
```

```
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1013
; LENGTH: 2454
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1013

Alignment Scores:
Pred. No.: 46 Length: 2454
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-138 (1-12) x US-09-489-039A-1013 (1-2454)

QY 4 ArgAsnAlaSerGlyLeuTyr 10
DB 726 CGAAACGCGCTGGACTATAT 746

RESULT 8
US-09-949-016-13593/c
; Sequence 13593, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13593
; LENGTH: 104077
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(104077)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13593

Alignment Scores:
Pred. No.: 1.79e+03 Length: 104077
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-138 (1-12) x US-09-949-016-13593 (1-104077)

QY 1 LeuGluTyrArgAsnAlaSer 7
DB 59717 TTGGAGTATAGGATGCTAGT 59697

RESULT 9
US-09-949-016-16514
; Sequence 16514, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

! TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

! FILE REFERENCE: CL001307
! CURRENT APPLICATION NUMBER: US/09/949,016
! CURRENT FILING DATE: 2000-04-14
! PRIOR APPLICATION NUMBER: 60/241,755
! PRIOR FILING DATE: 2000-10-20
! PRIOR APPLICATION NUMBER: 60/237,768
! PRIOR FILING DATE: 2000-10-03
! PRIOR APPLICATION NUMBER: 60/231,498
! PRIOR FILING DATE: 2000-09-08
! NUMBER OF SEQ ID NOS: 207012
! SOFTWARE: FastSeq for Windows Version 4.0
! SEQ ID NO 16514
! LENGTH: 160759
! TYPE: DNA
! ORGANISM: Human
! FEATURE:
! NAME/KEY: misc_feature
! LOCATION: (1)...(160759)
! OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16514

Alignment Scores:
Pred. No.: 2.75e+03 Length: 160759
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-138 (1-12) x US-09-949-016-16514 (1-160759)

Qy 5 AsnAlaSerGlyLeuTyrMet 11

Db 22644 AATGCTTCTGGTTATACATG 22664

RESULT 10

! Sequence 25272, Application US/09513999C
! Patent No. 6783961
! GENERAL INFORMATION:
! APPLICANT: Dumas Milne Edwards, J.B.
! APPLICANT: Duclert, A.
! APPLICANT: Giordano, J.Y.
! TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
! Patent No. 6783961
! FILE REFERENCE: 59, US2, REG
! CURRENT APPLICATION NUMBER: US/09/513,999C
! CURRENT FILING DATE: 2000-02-24
! PRIOR APPLICATION NUMBER: US 60/122,487
! PRIOR FILING DATE: 1999-02-26
! NUMBER OF SEQ ID NOS: 36681
! SOFTWARE: Patent.pm
! SEQ ID NO 25272
! LENGTH: 186
! TYPE: DNA
! ORGANISM: Homo sapiens
! FEATURE:
! NAME/KEY: misc_feature
! LOCATION: 104
! OTHER INFORMATION: k=g or t
US-09-513-999C-25272

Alignment Scores:
Pred. No.: 53.2 Length: 186
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-138 (1-12) x US-09-513-999C-25272 (1-186)

Qy 7 SerGlyLeuTyrMetVal 12

Db 84 AGTGGCCTCTATATGGTC 67

RESULT 11

! Sequence 12500, Application US/09248796A
! Patent No. 6747137
! GENERAL INFORMATION:
! APPLICANT: Keith Weinstock et al
! TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
! FILE REFERENCE: 107196.132
! CURRENT APPLICATION NUMBER: US/09/248,796A
! CURRENT FILING DATE: 1999-02-12
! PRIOR APPLICATION NUMBER: US 60/074,725
! PRIOR FILING DATE: 1998-02-13
! PRIOR APPLICATION NUMBER: US 60/096,409
! PRIOR FILING DATE: 1998-08-13
! NUMBER OF SEQ ID NOS: 28208
! SEQ ID NO 12500
! LENGTH: 210
! TYPE: DNA
! ORGANISM: Candida albicans
US-09-248-796A-12500/c

Alignment Scores:
Pred. No.: 59.9 Length: 210
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-138 (1-12) x US-09-248-796A-12500 (1-210)

Qy 7 SerGlyLeuTyrMetVal 12

Db 40 TCTGGTTTGTATATGGTG 23

RESULT 12

! Sequence 3, Application US/08917653
! Patent No. 6004751
! GENERAL INFORMATION:
! APPLICANT: Rosenfield, Robert L.
! TITLE OF INVENTION: IDENTIFICATION OF ACTIVATORS AND
! TITLE OF INVENTION: INHIBITORS OF SEBUM FORMATION
! NUMBER OF SEQUENCES: 4
! CORRESPONDENCE ADDRESS:
! ADDRESSEE: Arnold, White & Durkee
! STREET: P.O. Box 4433
! CITY: Houston
! STATE: Texas
! COUNTRY: U.S.
! ZIP: 77210
! COMPUTER READABLE FORM:
! MEDIUM TYPE: Floppy disk
! COMPUTER: IBM PC compatible
! OPERATING SYSTEM: PC-DOS/MS-DOS
! SOFTWARE: PatentIn Release #1.0, Version #1.30
! CURRENT APPLICATION DATA:
! APPLICATION NUMBER: US/08/917,653
! FILING DATE: Concurrently Herewith
! CLASSIFICATION: 435
! ATTORNEY/AGENT INFORMATION:
! NAME: Highlander, Steven L.
! REGISTRATION NUMBER: 37,642
! REFERENCE/DOCKET NUMBER: ARCD:216
! TELECOMMUNICATION INFORMATION:
! TELEPHONE: (512) 418-3000
! TELEFAX: (512) 474-7577
! INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 373 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-917-653-3

Alignment Scores:
Pred. No.: 105 Length: 373
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.00% Indels: 0
DB: 3 Gaps: 0

US-09-851-138C-138 (1-12) x US-08-917-653-3 (1-373)

QY 4 ArgAenAlaSerGlyLeu 9
Db 217 AGAATGCGAGTGGTCTT 200

RESULT 13

US-09-976-594-1081/c
Sequence 1081, Application US/09976594
Patent No. 6673549
GENERAL INFORMATION:
APPLICANT: Furness, Michael
TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
FILE REFERENCE: PA-0041 US
CURRENT APPLICATION NUMBER: US/09/976,594
CURRENT FILING DATE: 2001-10-12
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 1143
SOFTWARE: PERL Program
SEQ ID NO 1081
LENGTH: 431
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6673549 312986.1
US-09-976-594-1081

Alignment Scores:
Pred. No.: 121 Length: 431
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-138 (1-12) x US-09-976-594-1081 (1-431)

QY 4 ArgAenAlaSerGlyLeu 9
Db 393 AGAATGCATCTGGGCTC 376

RESULT 14

US-08-927-219-40
Sequence 40, Application US/08927219
Patent No. 6187533
GENERAL INFORMATION:
APPLICANT: Bell, Graeme I.
APPLICANT: Yamagata, Kazuya
APPLICANT: Oda, Naohisa
APPLICANT: Kaisaki, Pamela J.
APPLICANT: Furuta, Hiroto
APPLICANT: Horikawa, Yukio
APPLICANT: Menzel, Stephen
TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA

TITLE OF INVENTION: AND HNF-4ALPHA
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/08/927,219
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,679
FILING DATE: 30-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,056
FILING DATE: 02-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/025,719
FILING DATE: 10-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: ARCD:272
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 458 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: join(171..173, 177..265)
US-08-927-219-40

Alignment Scores:
Pred. No.: 128 Length: 458
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.00% Indels: 0
DB: 3 Gaps: 0

US-09-851-138C-138 (1-12) x US-08-927-219-40 (1-458)

QY 4 ArgAenAlaSerGlyLeu 9
Db 233 AGAATGCTTCGGGCTG 250

RESULT 15

US-09-902-540-8014
Sequence 8014, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkie, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883

; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 8014
; LENGTH: 474
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-8014

Alignment Scores:
Pred. No.: 133 Length: 474
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-138 (1-12) x US-09-902-540-8014 (1-474)

Qy 4 ArgAsnAlaSerGlyLeu 9
|||
Db 245 CGCAACGCCCTCTGGGCTT 262

Search completed: March 3, 2005, 22:05:12
Job time : 38.7385 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:43:48 ; Search time 756.185 Seconds
(without alignments)
604.047 Million cell updates/sec

Title: US-09-851-138c-138
Perfect score: 12
Sequence: 1 LEYRNAGLYMV 12

Scoring table: OLOGO?
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Word size: 1

Total number of hits satisfying chosen parameters: 68472171

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlp
-Oa/cpn2_1/USPTO spoel_p/US09851138/runat 28022005 120306 21476/app query.fasta_1.1123
-DB=EST-QPMT=fastap-SUFFIX=olog.rst-MINMATCH=0.1-LOOPCL=0-LOOPEXT=0
-UNITS=bits-START=1-END=1-MATRIX=oligo-TRANS=human40.coi-List=45
-DOCALIGN=200-THR SCORE=quality-THR_MIN=1-ALIGN=15-MODE=LOCAL-OUTFMT=ptc
-NORM=ext-HEAPSIZE=500-MINLEN=0-MAXLEN=2000000000
-NO WMAP-LARGEQUERY-NEG SCORES=0-WAIT DSPLOCK=100-LONGLOG
-DEV TIMEOUT=120-WARN TIMEOUT=30-THREADS=1-XGAPOP=60-DELOP=60-FCAPEXT=7-XGAPEXT=60-YGAPEXT=60-DELEXT=7

Database : EST:*

1:	gb_est1:*
2:	gb_est2:*
3:	gb_hic:*
4:	gb_est3:*
5:	gb_est4:*
6:	gb_est5:*
7:	gb_est6:*
8:	gb_g881:*
9:	gb_g882:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	8	66.7	235	2 AW607315	AW607315 QV4-HT046
C 2	7	58.3	137	1 AV624761	AV624761 AV624761
C 3	7	58.3	176	1 AV389042	AV389042 AV389042
C 4	7	58.3	201	9 CG595279	CG595279 OST254566
C 5	7	58.3	210	6 CD853797	CD853797 DH0AM222
C 6	7	58.3	271	5 BU646235	BU646235 1112048E1
C 7	7	58.3	324	7 CK371281	CK371281 zmrw005
C 8	7	58.3	340	7 CO275422	CO275422 EK117543
C 9	7	58.3	383	5 BU034395	BU034395 QHU3P10.Y

10	7	58.3	389	7	CO345371	CO345371	EP21921.3
11	7	58.3	397	2	BE417098	BE417098	MUG016.E1
C 12	7	58.3	405	7	CF078784	CF078784	QHK4A07.Y
C 13	7	58.3	410	2	BE725477	BE725477	894083F02
C 14	7	58.3	424	7	CO328360	CO328360	EP09052.3
15	7	58.3	428	8	AZ751482	AZ751482	RPCI-24-1
16	7	58.3	439	1	AL800562	AL800562	AL800562
C 17	7	58.3	439	2	BE211877	BE211877	894026G06
C 18	7	58.3	447	5	BU034024	BU034024	QHU2004.Y
C 19	7	58.3	450	4	BI724711	BI724711	1031074D0
C 20	7	58.3	450	7	CF078776	CF078776	QHK3P21.Y
C 21	7	58.3	456	2	AW159731	AW159731	2b06C07.X
22	7	58.3	456	9	CE106361	CE106361	tigr-g88-
23	7	58.3	458	8	AZ495426	AZ495426	1M0331B20
C 24	7	58.3	469	2	BE337091	BE337091	894043G09
C 25	7	58.3	475	1	AL927422	AL927422	AL927422
C 26	7	58.3	476	2	BE725544	BE725544	894084B06
C 27	7	58.3	492	7	CO280998	CO280998	EK155705.
C 28	7	58.3	498	4	BI964739	BI964739	1e55h10.Y
C 29	7	58.3	500	8	AQ776246	AQ776246	HS-3119.A
C 30	7	58.3	507	5	BU655694	BU655694	1112122F0
C 31	7	58.3	507	7	CN485704	CN485704	EST0350.P
C 32	7	58.3	529	8	AQ922245	AQ922245	RPCI-23-2
C 33	7	58.3	542	5	BQ818755	BQ818755	1030073A1
C 34	7	58.3	546	8	AZ408409	AZ408409	1M0179P03
C 35	7	58.3	548	7	CF088072	CF088072	QHM17N11.
C 36	7	58.3	552	8	AZ476881	AZ476881	1M0296L01
C 37	7	58.3	556	5	BP505099	BP505099	BP505099
C 38	7	58.3	574	4	BM000999	BM000999	1031092E1
C 39	7	58.3	595	5	BU650475	BU650475	1112087B0
C 40	7	58.3	604	9	FR0048327	FR0048327	AL604136
C 41	7	58.3	615	6	CD855728	CD855728	Fugu kubr
C 42	7	58.3	669	1	AL656707	AL656707	DH0AF232D
43	7	58.3	681	9	AG056465	AG056465	Pan trogl
44	7	58.3	686	9	CE097554	CE097554	tigr-g88-
45	7	58.3	697	2	BB199758	BB199758	BB199758

ALIGNMENTS

RESULT 1
AW607315/c 235 bp mRNA linear EST 23-MAR-2000
LOCUS QV4-HT0468-270100-093-a10 HT0468 Homo sapiens cDNA, mRNA sequence.
DEFINITION AW607315
ACCESSION AW607315
VERSION AW607315.1 GI:7311965
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 235)
HCGP <http://www.ludwig.org.br/ORESTES>.
The FAPESP/LICR Human Cancer Genome Project
Unpublished (1999)
CONTACT: Simpson A.J.G.
LABORATORY: Laboratory of Cancer Genetics
INSTITUTE: Ludwig Institute for Cancer Research
ADDRESS: Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
TEL: +55-11-2704922
FAX: +55-11-2707001
EMAIL: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV4&t2=QV4-HT0468-270100-093-a10&t3=2000-01-27&t4=1>)

Seq primer: puc 18 forward
High quality sequence start: 16
High quality sequence stop: 187.

FEATURES
Location/Qualifiers
1..235
/organism="Homo sapiens"

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0468"
/note="Organ: head neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN
Alignment Scores:
Pred. No.: 28 Length: 235
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 66.67% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-138 (1-12) x AW607315 (1-235)

Qy 2 GlnTyrArgAsnAlaSerGlyLeu 9
|||||
Db 160 GAATATAGGAACGCGCTGGACTA 137

RESULT 2
AV624761/c
LOCUS AV624761 137 bp mRNA linear EST 15-DEC-2000
DEFINITION Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
reinhartii cDNA clone LC082c08_r 5', mRNA sequence.
ACCESSION AV624761
VERSION AV624761.1 GI:10773938
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
1 (bases 1 to 137)
Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y. and Tabata, S.
Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
DNA Res. 7 (5), 305-307 (2000)
MEDLINE 20539644
PUBMED 11089912
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..137
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/db_xref="taxon:3055"
/clone="LC082c08_r"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"

ORIGIN
Alignment Scores:
Pred. No.: 214 Length: 137
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 0 Gaps: 0

US-09-851-138C-138 (1-12) x AV389042 (1-176)

Qy 6 AlaSerGlyLeuTyrMetVal 12
|||||
Db 82 GCTTCGTGCTTTATATAGGTT 62

RESULT 4
CG595279
LOCUS CG595279 201 bp mRNA linear GSS 02-OCT-2003
DEFINITION OST254566 Mus musculus 1298v/Ev Mus musculus cDNA clone OST254566,
mRNA sequence.
ACCESSION CG595279
VERSION CG595279.1 GI:37407342
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 201)

```


Research 6: 791-806."

ORIGIN

Alignment Scores:
 Pred. No.: 370 Length: 271
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 58.33% Indels: 0
 DB: 5 Gaps: 0

US-09-851-138C-138 (1-12) x BU646235 (1-271)

QY 4 ArgAsnAlaSerGlyLeuTyr 10

|||||
 30 AGGAATGCAAGCGCTTATAT 10

RESULT 7

CK371281 324 bp mRNA linear EST 23-DEC-2003
 DEFINITION zmrw005.0B10-006-b02.s0 zmrw005 Zea mays cDNA 5', mRNA sequence.

ACCESSION CK371281

VERSION CK371281.1 GI:40337211

KEYWORDS EST.

SOURCE Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 324)

Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredrickson,M.,
 Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A.,
 Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.

NSF Grant DBI-0211842; Functional Genomics of Root Growth and Root
 Signaling Under Drought

Unpublished (2003)

Contact: Hans Bohnert

University of Illinois, Urbana-Champaign

1201 West Gregory Drive, Urbana, IL 61801, USA

Tel: 217-265-5475

Fax: 217-333-5574

Email: bohnert@life.uiuc.edu

POLYA=NO.

FEATURES

source

1..324
 /organism="Zea mays"
 /mol_type="mRNA"
 /db_xref="taxon:4577"
 /clone_lib="zmrw005"
 /notes=" Library zmrw005 consists of the same cDNA material as library zmrw00 (described below) but was sequenced from the 5' prime end. The sequence identifier uses the 's0' suffix because the library tag was at the 3' prime end and thus not identified. Samples were collected in Robert E. Sharp's lab (University of Missouri-Columbia) to construct three normalized cDNA libraries. Dark-grown maize seedlings with primary roots 12-20 mm in length were transplanted to high (-0.03 MPa) or low water potential (-1.6 MPa) vermiculite, and harvested at 5 h and 48 h after transplanting. About 1,000 roots were used for each of the low water potential libraries (zmrw05 and zmrw48) while 500 roots were combined from each of the two time points at high water potential (zmrw00). Each root was divided into 4 segments (distances are from the junction of the root apex and root cap): segment 1, 0-3 mm plus the root cap; segment 2, 3-7 mm; segment 3, 7-12 mm; segment 4, 12-20 mm. (For details of conditions see (1) with nutrient modifications as in (2)). The three normalized cDNA libraries were constructed in the lab of Hans Bohnert (University of Illinois-UC). Total RNA was extracted by the 'hot Phenol' method (Plant Molecular Biology manual, DS: 1-13, 2nd ed., 1997). This method worked in

eliminating carbohydrate material present in the root tips. The integrity of the RNA was verified by denaturing agarose gels and spectrophotometry (ratio A260/280). Poly(A)⁺mRNA was isolated twice from total RNA using the Oligotex Direct mRNA kit (Qiagen). Poly(A)⁺ mRNA was converted to double-stranded cDNA and tagged by using modified Oligo(dT) primers. One of 4 sequence tags corresponding to a different segment of the root was added to the 3'-end of the modified Oligo(dT) primers, including a NotI site and used to reverse transcribe the segment-specific mRNAs into cDNAs. Each library contains all four tags. A suffix (s1, s2, s3, or s4) has been added to each sequence identifier to designate which region of the root (Root segment 1, 2, 3, or 4) the sequence was found in based on the identification of the tag. A suffix of s0 indicates that the sequence tag, and hence the source segment, could not be identified. The double stranded cDNAs were size-selected (>450 bp). Size selected cDNAs were adapted with EcoRI adaptors at both ends, and then digested with NotI. The cDNA was directionally cloned into EcoRI-NotI digested pBS II SK(+) phagemid vector (Stratagene) and electroporated into E.coli DH10B. The total number of white colony forming units (cfu) in the primary libraries before amplification was as follows: zmrw05: 3.37 x 10⁶; zmrw48: 4.87 x 10⁶; zmrw00: 3 x 10⁶. The background of empty clones was less than 1%. Inserts ranged from ~0.5kb to >2.5 kb, as determined by PCR. Plasmid DNA from the primary libraries then was converted to single-stranded circles and used as a template for PCR amplification using the T7 and T3 priming sites that flank the cloned cDNA inserts. The purified PCR products, representing the entire cDNA population cloned in each library, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30°C. Non-hybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B. The total number of clones with insert was: zmrw05: 2.0x10⁷; zmrw48: 4.2x10⁷; zmrw00: 1.1x10⁷. The background of empty clones was less than 2%. Insert size, determined by PCR of the entire library, ranged from 0.5kb to 2.5kb. (1) Sharp R E; Silk W K; Hsiao T C. Growth of the Maize Primary Root at Low Water Potentials I. Spatial Distribution of Expansive Growth. Plant Physiology (Rockville). 87(1): 1988. 50-57. (2) Spollen W G; LeNoble M E; Samuels T D; Bernstein N; Sharp R E. Abscissic acid accumulation maintains maize primary root elongation at low water potentials by restricting ethylene production. Plant Physiology (Rockville). 122(3). March, 2000. 967-976.

TAG ISSUE=Not found
 TAG_SEQ=Not found"

ORIGIN

Alignment Scores:

Pred. No.: 426 Length: 324
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 58.33% Indels: 0
 DB: 7 Gaps: 0

US-09-851-138C-138 (1-12) x CK371281 (1-324)

QY 3 TyrArgAsnAlaSerGlyLeu 9

|||||
 Db 277 TACAGGACGCCTCAGGACTC 297

RESULT 8

CO275422

LOCUS

DEFINITION

CO275422 340 bp mRNA linear EST 24-JUN-2004
 EK117543.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila

melanogaster cDNA clone EK117543 5, mRNA sequence.
 CO275422
 CO275422.1 GI:49196441
 EST.
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 340)
 Kopczynski, C., Platt, D., Campbell, J., Muzong, C., Laufer, A.,
 Peterson, E. and Swimmer, C.
 Exelixis FlyTag EST Project CK01 Library
 Unpublished (2004)
 COMMENT
 BDBG
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 Based upon one or more reads of this clone where vector sequence
 was present at both ends, this clone has been determined to contain
 contain a cDNA insert on the order of 600-1000 bases.
 Plate: EK.1175 row: D column: 7
 High quality sequence stop: 339.
 Location/Qualifiers
 1..340
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="EK117543"
 /note="Organ: mixed stage embryos, imaginal disks, and
 adult heads; Vector: pCDNA-SK+; Site 1: NotI; Site 2:
 XhoI; Random primed, normalized library from mixed stage
 embryos, imaginal disks, and adult heads."

Alignment Scores:
 Pred. No.: 443 Length: 340
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 58.33% Indels: 0
 DB: 7 Gaps: 0

US-09-851-138C-138 (1-12) x CO275422 (1-340)

Qy 4 ArgAnAlaSerGlyLeuTyr 10
 |||||
 Db 170 CQTACGCAAGCGGATTATAC 190
 |||||

RESULT 9
 BU034395/c
 LOCUS
 DEFINITION
 clone QHJ3P10.YG.ab1 QH EFGHJ sunflower RHA280 Helianthus annuus cDNA
 clone QHJ3P10, mRNA sequence.
 ACCESSION
 BU034395
 VERSION
 BU034395.1 GI:22469915
 KEYWORDS
 EST.
 SOURCE
 Helianthus annuus (common sunflower)
 ORGANISM
 Helianthus annuus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Asteroidae;
 Heliantheae; Helianthus.
 1 (bases 1 to 383)
 Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
 Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
 Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
 Lai, Z., Church, S., Jackson, L. and Bradford, K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
<http://compgenomics.ucdavis.edu/>

Unpublished (2002)
 Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Amundson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
 singleton, see <http://cgpdb.ucdavis.edu/> for details.
 Plate: QHJ3 row: P column: 10.
 Location/Qualifiers
 1..383
 /organism="Helianthus annuus"
 /mol_type="mRNA"
 /cultivar="RHA280"
 /db_xref="taxon:4232"
 /clone="QHJ3P10"
 /lab_hosts="E.coli"
 /note="Vector: pRCDNA-SK+; The library was constructed
 from 11 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at <http://cgpdb.ucdavis.edu/>
 TAG TISSUE=germinating seeds
 TAG_LIB=QH EFGHJ sunflower RHA280
 TAG_SEQ=TCTGTGCGGG"

Alignment Scores:
 Pred. No.: 487 Length: 383
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 58.33% Indels: 0
 DB: 5 Gaps: 0

US-09-851-138C-138 (1-12) x BU034395 (1-383)

Qy 1 LeuGluTyrArgAsnAlaSer 7
 |||||
 Db 261 CTAGAGTACAGAAATGCAAGC 241
 |||||

RESULT 10
 CO345371
 LOCUS
 DEFINITION
 melanogaster cDNA clone EP21921 3, mRNA sequence.
 ACCESSION
 CO345371
 VERSION
 CO345371.1 GI:49405678
 KEYWORDS
 EST.
 SOURCE
 Drosophila melanogaster (fruit fly)
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 389)
 Platt, D., Kopczynski, C., Muzong, C., Laufer, A., Leung, W.,
 Peterson, E. and Swimmer, C.
 Exelixis FlyTag EST Project CK02 Library
 Unpublished (2004)
 COMMENT
 BDBG
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 Plate: EP.219 row: B column: 9
 High quality sequence stop: 343.

JOURNAL
 COMMENT
 Unpublished (2002)
 Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Amundson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
 singleton, see <http://cgpdb.ucdavis.edu/> for details.
 Plate: QHJ3 row: P column: 10.
 Location/Qualifiers
 1..383
 /organism="Helianthus annuus"
 /mol_type="mRNA"
 /cultivar="RHA280"
 /db_xref="taxon:4232"
 /clone="QHJ3P10"
 /lab_hosts="E.coli"
 /note="Vector: pRCDNA-SK+; The library was constructed
 from 11 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at <http://cgpdb.ucdavis.edu/>
 TAG TISSUE=germinating seeds
 TAG_LIB=QH EFGHJ sunflower RHA280
 TAG_SEQ=TCTGTGCGGG"

Alignment Scores:
 Pred. No.: 487 Length: 383
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 58.33% Indels: 0
 DB: 5 Gaps: 0

US-09-851-138C-138 (1-12) x BU034395 (1-383)

Qy 1 LeuGluTyrArgAsnAlaSer 7
 |||||
 Db 261 CTAGAGTACAGAAATGCAAGC 241
 |||||

RESULT 10
 CO345371
 LOCUS
 DEFINITION
 melanogaster cDNA clone EP21921 3, mRNA sequence.
 ACCESSION
 CO345371
 VERSION
 CO345371.1 GI:49405678
 KEYWORDS
 EST.
 SOURCE
 Drosophila melanogaster (fruit fly)
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 389)
 Platt, D., Kopczynski, C., Muzong, C., Laufer, A., Leung, W.,
 Peterson, E. and Swimmer, C.
 Exelixis FlyTag EST Project CK02 Library
 Unpublished (2004)
 COMMENT
 BDBG
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 Plate: EP.219 row: B column: 9
 High quality sequence stop: 343.

FEATURES
source

Location/Qualifiers
1. .389
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="EF21921"
/clone_lib="Exelixis FlyTag CK02 pCDNA-SK+"
/notes="Organ: mixed stage embryos, imaginal disks, and adult heads; Vector: pCDNA-SK+; Site 1: NotI; Site 2: XhoI; Random primed, normalized library from mixed stage embryos, imaginal disks, and adult heads. Subset of Exelixis FlyTag CK01 clones sequenced from 3' end"

ORIGIN

Alignment Scores:
Pred. No.: 493 Length: 389
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 7 Gaps: 0

US-09-851-138C-138 (1-12) x C0345371 (1-389)

Qy 4 ArgAsnAlaSerGlyLeuTyr 10

Db 166 CGTAACGCAAGCGATTATAC 186

RESULT 11

BE417098
LOCUS BE417098 397 bp mRNA linear EST 24-JUL-2000
DEFINITION MUG016.E12R390620 ITEC MUG Wheat Spikelet Library Triticum aestivum
CDNA clone MUG016.E12, mRNA sequence.

ACCESSION BE417098

VERSION BE417098.1 GI:9414944

KEYWORDS EST.

SOURCE Triticum aestivum (bread wheat)

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.

1 (bases 1 to 397)

REFERENCE
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.
International Triticaceae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticeae Unpublished (2000)

TITLE Contact: Ogihara Y
Kihara Institute for Biological Research, Yokohama City University
Maioaka-cho 641-12, Totsuka-ku, Yokohama 244-0813, JAPAN
Tel: 81 45 820 1903
Fax: 81 45 820 1901
Email: ogihara@yokohama-cu.ac.jp
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.

FEATURES
source

Location/Qualifiers
1. .397
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Norin 26"
/db_xref="taxon:4565"
/clone="MUG016.E12"
/tissue_type="young spikelets"
/dev_stage="Feekes' scale 6-7"
/clone_lib="ITEC MUG Wheat Spikelet Library"
/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; M13 Reverse sequencing primer used. 1.2 Kbp average insert size."

ORIGIN

Alignment Scores:

Pred. No.: 501 Length: 397
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-138 (1-12) x BE417098 (1-397)

Qy 6 AlaSerGlyLeuTyrMetVal 12

Db 363 GCGAGTGGTCTTATATGCTT 383

RESULT 12

CF078784/c

DEFINITION

CF078784

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Helianthus paradoxus

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Asteraceae; Asteroideae; Heliantheae; Helianthus.

1 (bases 1 to 405)

REFERENCE

AUTHORS

Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L., Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z., Church,S., Jackson,L. and Bradford,K.

Lettuce and Sunflower ESTs from the Compositae Genome Project

http://compgenomics.ucdavis.edu/

Unpublished (2002)

COMMENT

Contact: Alexander Kozik [R.W.Michelmore]

Department of Vegetable Crops, R.W.Michelmore Lab

University of California at Davis (UCD)

Asmudson Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]

belongs to contig QH_CA_Contig4789, see http://cgdb.ucdavis.edu/

for details.

Plate: QHK4 row: A column: 07.

FEATURES

source

Location/Qualifiers

1. .405

/organism="Helianthus paradoxus"

/mol_type="mRNA"

/db_xref="taxon:73304"

/clone="QHKA07"

/lab_host="E.coli"

/clone_lib="QH K sunflower H.paradoxus"

/note="Vector: pBRCDNASFIAB; The library was constructed from four different sources (seedling, root, leaf and flower) of RNA from a single genotype. cDNAs were pooled and directionally cloned into a custom medium-copy vector. Details of library construction can be obtained at http://cgdb.ucdavis.edu/"

ORIGIN

Alignment Scores:

Pred. No.: 509 Length: 405
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.33% Indels: 0
DB: 7 Gaps: 0

US-09-851-138C-138 (1-12) x CF078784 (1-405)

QY 1 LeuGlutTyArgAsnAlaSer 7
 Db 274 CTAGAGTACAGAAATGCAAGC 254

RESULT 13
 BE725477/c
 LOCUS 894083F02.y1 C. reinhardtii CC-1690, linear EST 14-SEP-2000
 DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BE725477
 VERSION BE725477.1 GI:10126773
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii
 ORGANISM Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 410)
 AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,
 McDermott, J. P., Sliflow, C., Stern, D., and Surzycki, R.
 TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants; project phase 2
 JOURNAL Unpublished (2000)
 COMMENT Contact: Charles Hauser
 DCMB Box 91000
 Duke University
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chause@duke.edu.

FEATURES
 source
 1..410
 Location/Qualifiers
 /organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
 II"
 /notes="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
 XhoI; This library, constructed by John Davies and Jeffrey
 McDermott, combines cDNAs from CC-1690 cells grown to
 mid-log phase in TAP (acetate-containing) medium in the
 light, TAP medium in the dark, HS (minimal) medium in
 ambient levels of CO₂ and HS medium bubbled with 5% CO₂.
 PolyA mRNA was purified from each sample, pooled and cDNA
 synthesized. The cDNA was directionally cloned into lambda
 Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
 pBluescript II SK- plasmids were excised from the lambda
 Zap clones by superinfection with ExAssist (Stratagene)
 phage. The library was normalized using method 4 described
 in Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN
 Alignment Scores:
 Pred. No.: 514 Length: 410
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 58.33% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-138 (1-12) x BE725477 (1-410)

QY 6 AlaSerGlyLeuTyMetVal 12
 Db 71 GCTTCGGCTTATATAGGTT 51

RESULT 14
 CO328360/c
 LOCUS 424 bp mRNA linear EST 28-JUN-2004
 DEFINITION EP09052.3prime Exelixis FlyTag CK02 pCDNA-SK+ Drosophila
 melanogaster cDNA clone EP09052 3, mRNA sequence.

ACCESSION CO328360

VERSION CO328360.1 GI:49386794
 KEYWORDS EST.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 424)
 AUTHORS Platt, D., Koczynski, C., Muzong, C., Laufer, A., Leung, W.,
 Peterson, E., and Swimmer, C.
 TITLE Exelixis FlyTag EST Project CK02 Library
 JOURNAL Unpublished (2004)
 COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
 Plate: EP.90 row: E column: 4
 High quality sequence stop: 309.

FEATURES
 source
 1..424
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone_lib="EP09052"
 /note="Organ: mixed stage embryos, imaginal disks, and
 adult heads; Vector: pCDNA-SK+; Site 1: NotI; Site 2:
 XhoI; Random primed, normalized library from mixed stage
 embryos, imaginal disks, and adult heads. Subset of
 Exelixis FlyTag CK01 clones sequenced from 3' end"

ORIGIN
 Alignment Scores:
 Pred. No.: 528 Length: 424
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 58.33% Indels: 0
 DB: 7 Gaps: 0

US-09-851-138C-138 (1-12) x CO328360 (1-424)

QY 4 ArgAsnAlaSerGlyLeuTy 10
 Db 34 CGTACGCGACGGGATATAC 14

RESULT 15
 AZ751482
 LOCUS 428 bp DNA linear GSS 25-JAN-2001
 DEFINITION RPCI-24-102C4.TJ RPCI-24 Mus musculus genomic clone RPCI-24-102C4,
 genomic survey sequence.

ACCESSION AZ751482
 VERSION AZ751482.1 GI:12536641
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 428)

REFERENCE 1 (bases 1 to 428)
 AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
 Tsagaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
 Russell, D., de Jong, P., and Fraser, C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-24
 JOURNAL Unpublished (1999)
 COMMENT Other GSSs: RPCI-24-102C4.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdj@tigr.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
 Plate: 102 row: C column: 4
 Seq primer: SP6
 Class: BAC ends.

FEATURES

source
 Location/Qualifiers
 1..428
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-102C4"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /clone_lib="RPCI-24"
 /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamH1 sites using MboI partially digested male C57BL/6J DNA."
 BAC ends.

ORIGIN

Alignment Scores:
 Pred. No.: 532 Length: 428
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 58.33% Indels: 0
 DB: 8 Gaps: 0

US-09-851-138C-138 (1-12) x AZ751482 (1-428)

QY 4 ArgAsnAlaSerGlyLeuTyr 10
 Db 196 AGAAATGCTTCTGGCCITAT 216

Search completed: March 3, 2005, 21:58:12
 Job time : 761.185 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 14:30:42 ; Search time 78.4667 Seconds
(without alignments)
829.870 Million cell updates/sec

Title: US-09-851-138C-155
Perfect score: 11
Sequence: 1 VYEAGDIILHL 11

Scoring table: OXIGOO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8763375

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cpn2.1/uspto.spool.p/US09851138/runat.28022005.120306.21457/app.query.fasta_1.1123
-DB=N_Geneseq.16Dec04 -QFMT=fastap -SUFFIX=olg.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=olg -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09851138 @CGN 1 1 1418 @runat.28022005.120306.21457 -NCPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq.16Dec04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	100.0	447	AAT27962	Aat27962 Hepatitis
C 2	7	63.6	446	ABV14335	Abv14335 Human pro
C 3	7	63.6	489	ABV44251	Abv44251 Human pro
C 4	7	63.6	489	ABV35422	Abv35422 Human pro
C 5	7	63.6	606	ADS57861	Ads57861 Bacterial

C 6	7	63.6	620	13	ADR59409	Adt59409 Cotton cD
C 7	7	63.6	774	8	ABT21005	Abt21005 Aspergill
C 8	7	63.6	774	8	ABT19185	Abt19185 Aspergill
C 9	7	63.6	977	8	ABT18591	Abt18591 Aspergill
C 10	7	63.6	977	8	ABT20407	Abt20407 Aspergill
C 11	7	63.6	1304	6	ABK65231	Abk65231 Arabidops
C 12	7	63.6	1304	10	ADC46624	Adc46624 Thalectes
C 13	7	63.6	1304	10	ADD31048	Ad31048 Plant yie
C 14	7	63.6	1304	10	ADE31460	Ade31460 Plant yie
C 15	7	63.6	1304	12	ADI41750	Adi41750 Plant tra
C 16	7	63.6	1304	12	ADI61338	Adi61338 cDNA enco
C 17	7	63.6	1304	12	ADO02286	Ado02286 Thalectes
C 18	7	63.6	1504	3	AAK38952	Aac38952 Arabidops
C 19	7	63.6	2976	8	ABT17997	Abt17997 Aspergill
C 20	7	63.6	2976	8	ABT19811	Abt19811 Aspergill
C 21	7	63.6	14301	4	ABL02084	Abi02084 Drosophil
C 22	7	63.6	137870	10	ADG89426	Adg89426 Human mat
C 23	7	63.6	165199	6	ABK83460	Abk83460 Human cD
C 24	7	63.6	196063	13	ABD33462	Abd33462 Human can
C 25	6	54.5	20	10	ABZ87368	Abz87368 Human oli
C 26	6	54.5	20	11	ABD23598	Abd23598 Human myo
C 27	6	54.5	25	9	ACI65552	Act65552 Human mic
C 28	6	54.5	27	5	AAF54973	Aaf54973 PCR prime
C 29	6	54.5	27	6	ABS64076	Ab64076 CGMW 129
C 30	6	54.5	60	2	AZ19673	Aaz19673 Complemen
C 31	6	54.5	74	2	AAV99400	Aav99400 Oligonuc
C 32	6	54.5	190	6	ABN19510	Abn19510 Human ORF
C 33	6	54.5	207	10	ABX06446	Abx06446 S. pneumo
C 34	6	54.5	207	12	ADM91893	Adm91893 S. pneumo
C 35	6	54.5	211	5	AAS77529	Aas77529 DNA encod
C 36	6	54.5	213	13	ADR91427	Adr91427 Novel S.
C 37	6	54.5	273	2	AAQ76615	Aaq76615 Human gen
C 38	6	54.5	308	10	ADB56588	Adb56588 Toxicity-
C 39	6	54.5	308	10	ABT41073	Abt41073 Toxicity
C 40	6	54.5	313	3	AACT4886	Aac74886 Human ORF
C 41	6	54.5	313	6	ABN79556	Abn79556 Human ORF
C 42	6	54.5	324	6	ABK79113	Abk79113 Bacillus
C 43	6	54.5	350	3	ACG01412	Aac01412 Human sec
C 44	6	54.5	357	12	ADJ39337	Adj39337 Plant cDN
C 45	6	54.5	371	5	AAS68585	Aas68585 DNA encod

ALIGNMENTS

RESULT 1
AAT27962
ID AAT27962 standard; DNA; 447 BP.
XX
AC AAT27962;
XX
DT 11-MAR-1997 (first entry)
XX
DE Hepatitis C virus type 10a isolate NN98 bases 478-925.
XX
KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
XX PCR; primer; probe; antibody; infection; ss.
OS Hepatitis C virus.
XX
PN WO9613590-A2.
XX
PD 09-MAY-1996.
XX
PF 23-OCT-1995; 95WO-EP004155.
PR 21-OCT-1994; 94EP-00870166.
PR 28-JUN-1995; 95EP-00870076.
XX
(INNO-) INNOGENETICS NV.
PA
XX
PI Maertens G, Stuyver L;
XX
DR WPI; 1996-251460/25.

DR P-PSDB; AAR96551.
 XX Hepatitis C virus poly:nucleic acid unique to unidentified sub-type -
 PT used to develop probes and primers for new subtypes and vaccines to
 PT prevent and treat infection.
 XX Claim 6; Fig 3; 150pp; English.
 XX The sequences AAT27937-T27989 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f, 4a-
 CC j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l, 3g,
 CC 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
 CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the genome.
 CC This sequence represents nucleotides 478-925 from the HCV type 10a
 CC isolate NB98. The new HCV types were isolated from patients with chronic
 CC HCV from the Benelux countries, France, Cameroon and Vietnam, because of
 CC their aberrant reactivities. The RNA was extracted, cDNA synthesised and
 CC PCR amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions
 CC were sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences. The
 CC sequences were used to generate the peptides AAR96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect anti
 CC -HCV antibodies, for HCV typing or to prevent HCV infections
 XX SQ Sequence 447 BP; 82 A; 130 C; 114 G; 118 T; 0 U; 3 Other;
 Alignment Scores:
 Pred. No.: 0.00155 Length: 447
 Score: 11.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0
 US-09-851-138C-155 (1-11) x AAT27962 (1-447)
 QY 1 ValTyrGluAlaGlyAspIleLeuHisLeu 11
 DB 160 GTGTATGAGCGGGGATATTATCTCCACTTA 192
 RESULT 2
 ABV14335/C
 ID ABV14335 standard; cDNA; 446 BP.
 XX AC ABV14335;
 XX DT 13-SEP-2002 (first entry)
 XX DE Human prostate expression marker cDNA 14326.
 XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX OS Homo sapiens.
 XX PN WO200160860-A2.
 XX PD 23-AUG-2001.
 XX PF 20-FEB-2001; 2001WO-US005171.
 XX PR 17-FEB-2000; 2000US-0183319P.
 XX PR 16-MAR-2000; 2000US-0189862P.
 XX PR 25-MAY-2000; 2000US-0207454P.
 XX PR 09-JUN-2000; 2000US-0211314P.
 XX PR 18-JUL-2000; 2000US-0219007P.
 XX PR 13-DEC-2000; 2000US-0255281P.
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX Schlegel R, Endege WO, Monahan JE;
 XX WPI; 2001-662795/76.
 XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 XX

DR WPI; 2001-662795/76.
 XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX Claim 1; Page 2392; 11750pp; English.
 XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX SQ Sequence 446 BP; 87 A; 86 C; 81 G; 192 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 42.8 Length: 446
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 5 Gaps: 0
 US-09-851-138C-155 (1-11) x ABV14335 (1-446)
 QY 1 ValTyrGluAlaGlyAspIle 7
 DB 437 GTCTATGAAGCAGCGGATATT 417
 RESULT 3
 ABV4251/C
 ID ABV4251 standard; cDNA; 489 BP.
 XX AC ABV4251;
 XX DT 16-SEP-2002 (first entry)
 XX DE Human prostate expression marker cDNA 4242.
 XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX OS Homo sapiens.
 XX PN WO200160860-A2.
 XX PD 23-AUG-2001.
 XX PF 20-FEB-2001; 2001WO-US005171.
 XX PR 17-FEB-2000; 2000US-0183319P.
 XX PR 16-MAR-2000; 2000US-0189862P.
 XX PR 25-MAY-2000; 2000US-0207454P.
 XX PR 09-JUN-2000; 2000US-0211314P.
 XX PR 18-JUL-2000; 2000US-0219007P.
 XX PR 13-DEC-2000; 2000US-0255281P.
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX Schlegel R, Endege WO, Monahan JE;
 XX WPI; 2001-662795/76.
 XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 XX

PT for detecting presence of prostate cancer, stage of prostate cancer.
PS Claim 1; Page 8788; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 489 BP; 94 A; 100 C; 96 G; 199 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 46.6 Length: 489
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 63.64% Indels: 0
DB: 5 Gaps: 0
US-09-851-138C-155 (1-11) x ABV44251 (1-489)
QY 1 ValTyrGluAlaGlyAspIle 7
DB 480 GTCTATGAAGCAGGGGATATT 460
RESULT 4
ABV35422/c
ID ABV35422 standard; cDNA; 489 BP.
XX
AC ABV35422;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 35413.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 7379; 11750pp; English.
XX

CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 489 BP; 94 A; 100 C; 96 G; 199 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 46.6 Length: 489
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 63.64% Indels: 0
DB: 5 Gaps: 0
US-09-851-138C-155 (1-11) x ABV35422 (1-489)
QY 1 ValTyrGluAlaGlyAspIle 7
DB 480 GTCTATGAAGCAGGGGATATT 460
RESULT 5
ADS57861/c
ID ADS57861 standard; cDNA; 606 BP.
XX
AC ADS57861;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polynucleotide #9848.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomanan;
KW bacterial polynucleotide; gene; ss.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 33535; 122pp; English.
XX

CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polynucleotide used in
 CC the scope of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.

SQ Sequence 606 BP; 149 A; 130 C; 185 G; 142 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 56.8 Length: 606
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 13 Gaps: 0

US-09-851-138C-155 (1-11) x ADS57861 (1-606)

Qy 5 GlyAspIleIleuHisLeu 11

Db 84 GGGGACATTATTCGCATCTT 64

RESULT 6

ID ADR59409/c

AD ADR59409 standard; cDNA; 620 BP.

AC ADR59409;

DT 02-DEC-2004 (first entry)

DE Cotton cDNA sequence, SEQ ID 190.

XX Cotton; ss; plant; cold tolerance; growth rate; cell cycle pathway;
 KW drought tolerance; plant disease resistance; galactomannan; lignin;
 KW plant growth regulator; heat tolerance; herbicide tolerance;
 KW homologous recombination; extreme osmotic condition tolerance;
 KW pathogen resistance; pest resistance; yield; photosynthesis; seed oil;
 KW stress resistance.

OS Gossypium hirsutum.

PN US2004181830-A1.

PD 16-SEP-2004.

PF 29-JAN-2004; 2004US-00767795.

PR 07-MAY-2001; 2001US-00849529.

PR 12-DEC-2001; 2001US-00021323.

PA (KOVA/) KOVALIC D K.

PA (ZHOU/) ZHOU Y.

PA (CAOY/) CAO Y.

XX

PI Kovalic DK, Zhou Y, Cao Y;

XX WPI; 2004-667718/65.

XX New recombinant nucleic acid molecules and polypeptides from Gossypium

PT hirsutum, useful for producing plants with improved biological

PT characteristics (e.g. improved plant cold or drought tolerance).

XX Claim 1; SEQ ID NO 190; 14pp; English.

XX The invention relates to a recombinant polynucleotide comprising any of

CC the 58798 Cotton plant cDNA sequences mentioned in the specification.

CC Also a recombinant polypeptide comprising any of the 58798 amino acid

CC sequences mentioned in the specification and producing a plant having an

CC improved property. Producing a plant having an improved property

CC comprises transforming a plant with a recombinant construct comprising a

CC promoter region functional in a plant cell operably joined to a

CC polynucleotide comprising a coding sequence for a polypeptide associated

CC with the property, and growing the transformed plant. The polypeptide is

CC useful for improving plant cold tolerance, manipulating growth rate in

CC plant cells by modification of the cell cycle pathway, improving plant

CC drought tolerance, providing increased resistance to plant disease,

CC producing galactomannan (or lignin or plant growth regulators), improving

CC plant heat tolerance, improving plant tolerance to herbicides, increasing

CC the rate of homologous recombination in plants, improving plant tolerance

CC to extreme osmotic conditions or to pathogens or pests, improving yield

CC by modification of photosynthesis, modifying seed oil or protein yield

CC and/or content, improving yield by modification of carbohydrate, nitrogen

CC or phosphorus use and/or uptake, or improving yield by providing improved

CC plant growth and development under at least one stress condition. The

CC polynucleotide and polypeptide may also be used in recombinant DNA

CC constructs, in physical arrays of molecules, as plant breeding markers,

CC or in computer-based storage and analysis systems. The present sequence

CC is a Cotton plant cDNA of the invention. NOTE: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from USPTO at

CC seqdata.uspto.gov/sequence.html?DocID=20040181830. However only 6585

CC polynucleotide sequences were available, the remaining 52213

CC polynucleotides and all 58798 protein sequences were not present.

SQ Sequence 620 BP; 192 A; 130 C; 144 G; 154 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 58 Length: 620
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 13 Gaps: 0

US-09-851-138C-155 (1-11) x ADR59409 (1-620)

Qy 3 GluAlaGlyAspIleIleu 9

Db 482 GAAGCAGGAGATATTATCTT 462

RESULT 7

ABT21005/c

ID ABT21005 standard; DNA; 774 BP.

XX AC ABT21005;

XX DT 16-APR-2003 (first entry)

XX DE Aspergillus fumigatus essential gene #3363.

XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;

XX cancer; contamination; biofilm; antibody; immune response; ds.

XX OS Aspergillus fumigatus.

XX WO200286090-A2.

XX

PD 31-OCT-2002.
 XX
 PF 23-APR-2002; 2002WO-US013142.
 XX
 PR 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287066P.
 PR 05-JUN-2001; 2001US-0295890P.
 PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.
 DR
 XX New purified or isolated nucleic acids of essential genes of *Aspergillus*
 PT fumigatus, useful for treating or preventing infections by *A. fumigatus*,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX
 PS Disclosure; Page; 175pp; English.
 XX
 CC The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as *A. fumigatus*, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by *A. fumigatus*, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising *A. fumigatus*. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of *A.*
 CC *fumigatus* to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of *Aspergillus fumigatus* of the invention
 XX
 SQ Sequence 774 BP; 189 A; 211 C; 217 G; 157 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 71.2 Length: 774
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 8 Gaps: 0
 US-09-851-138C-155 (1-11) x ABT21005 (1-774)
 QY 4 AlaGlyAspIleLeuHis 10
 DB 591 GCCGGTGACATCATCCTTCAT 571
 RESULT 8
 ABT19185/c
 ID ABT19185 standard; DNA; 774 BP.
 XX
 AC ABT19185;
 XX
 XX
 DT 16-APR-2003 (first entry)
 XX

DE *Aspergillus fumigatus* essential gene #1543.
 XX
 KW Fungicide; cytostatic; essential gene; *Aspergillus fumigatus*; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.
 XX
 OS *Aspergillus fumigatus*.
 XX
 PN WO200286090-A2.
 PD 31-OCT-2002.
 XX
 XX 23-APR-2002; 2002WO-US013142.
 PF
 XX 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287066P.
 PR 05-JUN-2001; 2001US-0295890P.
 PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.
 DR
 XX New purified or isolated nucleic acids of essential genes of *Aspergillus*
 PT fumigatus, useful for treating or preventing infections by *A. fumigatus*,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX
 PS Disclosure; Page; 175pp; English.
 XX
 CC The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as *A. fumigatus*, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by *A. fumigatus*, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising *A. fumigatus*. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of *A.*
 CC *fumigatus* to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of *Aspergillus fumigatus* of the invention
 XX
 SQ Sequence 774 BP; 189 A; 211 C; 217 G; 157 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 71.2 Length: 774
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 8 Gaps: 0
 US-09-851-138C-155 (1-11) x ABT19185 (1-774)
 QY 4 AlaGlyAspIleLeuHis 10
 DB 591 GCCGGTGACATCATCCTTCAT 571

RESULT 9
 ABT18591/C
 ID ABT18591 standard; DNA; 977 BP.
 XX
 AC ABT18591;
 XX
 DT 16-APR-2003 (first entry)
 XX
 DE Aspergillus fumigatus essential gene #949.
 XX
 KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.
 XX
 OS Aspergillus fumigatus.
 XX
 PN WO200286090-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 23-APR-2002; 2002WO-US013142.
 XX
 PR 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287066P.
 PR 05-JUN-2001; 2001US-0295890P.
 PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX
 DR WPI; 2003-093124/08.
 XX
 PT New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX
 PS Disclosure; Page; 175pp; English.
 XX
 CC The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention
 XX
 SQ Sequence 977 BP; 234 A; 251 C; 269 G; 223 T; 0 U; 0 Other;

Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 8 Gaps: 0
 US-09-851-138C-155 (1-11) x ABT18591 (1-977)
 QY 4 AlaGlyAspIleIleLeuHis 10
 DB 794 GCGGGTGACATCATCCTTCAT 774
 RESULT 10
 ABT20407/C
 ID ABT20407 standard; DNA; 977 BP.
 XX
 AC ABT20407;
 XX
 DT 16-APR-2003 (first entry)
 XX
 DE Aspergillus fumigatus essential gene #2765.
 XX
 KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.
 XX
 OS Aspergillus fumigatus.
 XX
 PN WO200286090-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 23-APR-2002; 2002WO-US013142.
 XX
 PR 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287066P.
 PR 05-JUN-2001; 2001US-0295890P.
 PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX
 DR WPI; 2003-093124/08.
 XX
 PT New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX
 PS Disclosure; Page; 175pp; English.
 XX
 CC The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention
 XX
 SQ Sequence 977 BP; 234 A; 251 C; 269 G; 223 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 88.2 Length: 977
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0

CC factors. This polynucleotide sequence represents one of the essential
 CC genes of *Aspergillus fumigatus* of the invention

SQ Sequence 977 BP; 234 A; 251 C; 269 G; 223 T; 0 U; 0 Other;

Alignment Scores: 88.2 Length: 977
 Pred. No.: 7.00 Matches: 7
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 100.00% Indels: 0
 Query Match: 63.64% Gaps: 0
 DB: 8

US-09-851-138C-155 (1-11) x ABT20407 (1-977)

QY 4 AlaGlyAspIleLeuHis 10

Db 794 GCCGGTGACATCATCTTCAT 774

RESULT 11

ID ABK65231/c

AC ABK65231 standard; cDNA; 1304 BP.

XX

XX

DT 02-JUL-2002 (first entry)

DE Arabidopsis cDNA encoding a transcription factor #83.

KW Plant; ss; gene; transcription factor; transgenic; agriculture;
 KW metabolic chemical; environmental stress; drought;
 KW microbial disease resistance; herbicide resistance; seed yield;
 KW fruit yield; growth rate; leaf senescence; flower senescence.

OS Arabidopsis thaliana.

XX WO200215675-A1.

XX 28-FEB-2002.

XX 22-AUG-2001; 2001WO-US026189.

XX 22-AUG-2000; 2000US-0227439P.

PR 16-NOV-2000; 2000US-00713994.

PR 18-APR-2001; 2001US-00837944.

XX (MEND-) MENDEL BIOTECHNOLOGY INC.

PA (PILG/) PILGRIM M.

PA (CREE/) CREELMAN R.

PA (DUBE/) DUBELL A J.

PA (HEAR/) HEARD J.

PA (JIAN/) JIANG C.

PA (KEDD/) KEDDIE J.

PA (ADAM/) ADAM L.

PA (RATC/) RATCLIFF O.

PA (REUB/) REUBER J L.

PA (RIEC/) RIECHMANN J L.

PA (YUGG/) YU G.

PA (PINE/) PINEDA O.

XX Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddle J;

PI Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;

XX WPI: 2002-292022/33.

DR P-PSDB; AAU93045.

XX An isolated or recombinant polynucleotide used to produce a transgenic

PT plant.

XX Claim 4; Page 367-369; 941pp; English.

XX The invention relates to 1 of 232 isolated or recombinant polynucleotides

CC encoding an Arabidopsis thaliana transcription factor, their variants,

CC complements, fragments, or related polynucleotide with 31% to 95%
 CC sequence identity, where the plant possesses an altered trait as compared
 CC to a wild-type or reference plant, or the plant exhibits an altered
 CC phenotype as compared to a wild-type or reference plant, or the plant
 CC exhibits ectopic expression or altered expression of one or more genes
 CC associated with a plant trait as compared to a wild plant. Also included
 CC are a transgenic plant comprising the polynucleotides, a computer
 CC readable medium having stored sequence information, and identifying a
 CC homologue sequence from a database comprising a plurality of known plant
 CC sequences comprising inputting sequence information selected from one of
 CC 464 fully defined sequences given in the specification. The isolated or
 CC recombinant polynucleotide is used for producing a plant having a
 CC modified trait, the method comprising selecting a polynucleotide that
 CC encodes a polypeptide or an antisense nucleic acid, inserting the
 CC polynucleotide or antisense nucleic acid into an expression vector,
 CC introducing the vector into a plant or a cell of a plant to overexpress
 CC the polypeptide or antisense nucleic acid, thereby producing a modified
 CC plant, and selecting for a modified trait (e.g. increased production of
 CC agriculturally useful proteins or metabolic chemicals, pest tolerance,
 CC environmental stress response (e.g. drought), microbial disease
 CC resistance, herbicide resistance, seed and fruit yield, growth rate, leaf
 CC and flower senescence and many other traits listed in the specification).
 CC The present sequence is one of the 232 polynucleotides encoding an A.
 CC thaliana transcription factor

SQ Sequence 1304 BP; 391 A; 322 C; 238 G; 353 T; 0 U; 0 Other;

Alignment Scores: 115 Length: 1304
 Pred. No.: 7.00 Matches: 7
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 100.00% Indels: 0
 Query Match: 63.64% Gaps: 0
 DB: 6

US-09-851-138C-155 (1-11) x ABK65231 (1-1304)

QY 5 GlyAspIleLeuHisLeu 11

Db 1067 GGAGACATTATTCATCTT 1047

RESULT 12

ADC46624/c

ID ADC46624 standard; DNA; 1304 BP.

XX

AC ADC46624;

XX

DT 18-DEC-2003 (first entry)

XX Thalecress transcription factor-like DNA G1052.

DE Thalecress; transcription factor-like protein; ds; seed trait;
 KW transgenic; gene; plant size; stress tolerance; yield;
 KW disease resistance; plant.

OS Arabidopsis thaliana.

XX US2003093837-A1.

XX 15-MAY-2003.

XX 01-NOV-2002; 2002US-00286264.

XX 23-MAR-1999; 99US-0125814P.

XX 22-MAR-2000; 2000US-00533030.

XX (KEDD/) KEDDIE J.

PA (RIEC/) RIECHMANN J L.

PA (RATC/) RATCLIFFE O.

PA (ZHAN/) ZHANG J.

PA (JIAN/) JIANG C.

PA (PINE/) PINEDA O.

PA (HEAR/) HEARD J.

PA (YUGG/) YU G.
PA (ADAM/) ADAM L.
PA (BROU/) BROUN P.
PA (REUB/) REUBER L.
PA (PILG/) PILGRIM M.
PA (SAMA/) SAMAHA R.
XX
PI Keddie J, Riechmann JL, Ratcliffe O, Zhang J, Jiang C, Pineda O;
PI Heard J, Yu G, Adam L, Broun P, Reuber L, Pilgrim M, Samaha R;
XX
XX WPI; 2003-765498/72.
XX P-PSDB; ADC46625.
XX
XX Novel transgenic plant having recombinant polynucleotide encoding
PT polypeptide that alters trait of transgenic plant when compared with same
PT trait of another plant lacking recombinant polynucleotide.
XX
XX Disclosure; SEQ ID NO 23; 165pp; English.
XX
XX The invention relates to a transgenic plant having recombinant
CC polynucleotide (II) encoding polypeptide comprising at least 6
CC consecutive amino acids of a sequence chosen from the protein sequence
CC appearing as ADC46603 - ADC46749 (every second sequence), where
CC recombinant polynucleotide alters a trait of the seed transgenic plant
CC when compared with same trait of another plant lacking recombinant
CC polynucleotide. The proteins are transcription factor-like proteins. Also
CC included are altering (M1) a trait associated with seed (comprising:
CC transforming a plant with (II); selecting the transformed plants; and
CC identifying a transformed plant with seed having altered trait), altering
CC (M2) the expression levels of at least one gene of a plant (involving
CC transforming the plant with (II) and selecting the transformed plant),
CC altering (M3) a trait associated with a plant's seed (comprising:
CC transforming the plant with a recombinant polynucleotide comprising a
CC nucleotide sequence comprising least 18 consecutive nucleotides of a
CC sequence appearing as ADC46750 - ADC46766 and selecting the transformed
CC plant) altering (M4) a plant's trait (involving providing a database
CC sequence, comparing the database sequence with a polypeptide or a
CC polynucleotide chosen as detailed above, selecting a database sequence
CC that needs selected sequence criteria and transforming a database
CC sequence in the plant) and altering a plant's trait (involving providing
CC a test polynucleotide, hybridising the test polynucleotide with a
CC polynucleotide detailed above at low stringency and transforming the
CC hybridising test polynucleotide in a plant to alter a trait of the
CC plant). The method (M1) is useful for altering a trait associated with
CC seed. The method (M2) is useful for altering the expressing levels of at
CC least one gene of a plant. The method (M3) is useful for altering a trait
CC associated with a plant's seed. The method (M4) is useful for altering a
CC plant's trait. The method (M4) is useful for altering a plant's trait
CC such as seed or plant size, stress tolerance, yield or disease
CC resistance. The present sequence encodes a transcription factor-like
CC protein/seed trait altering protein of the invention.
XX
SQ Sequence 1304 BP; 391 A; 322 C; 238 G; 353 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 115 Length: 1304
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 63.64% Indels: 0
DB: 10 Gaps: 0

US-09-851-138C-155 (1-11) x ADC46624 (1-1304)

Qy 5 GlyAspIleIleLeuHisLeu 11
|||||
Db 1067 GGAGACATTATTCCTCATCTT 1047

RESULT 13
ADD31048/C
ID ADD31048 standard; cDNA; 1304 BP.
XX
AC ADD31048;

XX 15-JAN-2004 (first entry)
XX Plant yield-related polynucleotide clone G1052.
DE
XX ds; transcription factor; transgenic plant; growth rate; senescence;
KW seed germination rate; plant vigor; seedling vigor.
XX Arabidopsis thaliana.
OS
XX WO2003013227-A2.
PN
XX 20-FEB-2003.
PD
XX 09-AUG-2002; 2002WO-US025805.
PF
XX 09-AUG-2001; 2001US-0310847P.
PR 19-NOV-2001; 2001US-0336049P.
PR 11-DEC-2001; 2001US-0338692P.
PR 14-JUN-2002; 2002US-00171468.
XX (MEND-) MENDEL BIOTECHNOLOGY INC.
PA Ratcliffe O, Riechmann JL, Adam LJ, Dubell AT, Heard JE;
PI Pilgrim ML, Jiang C, Reuber TL, Creelman RA, Pineda O, Yu G;
PI Broun PE;
XX WPI; 2003-248221/24.
DR P-PSDB; ADD31049.
DR
XX New plant transcription factor polynucleotides and polypeptides, useful
PT in producing transgenic plants with commercially valuable properties,
PT such as an alteration in a plant growth characteristic, e.g. growth rate
PT or apomixis.
XX
XX Disclosure; SEQ ID NO 1077; 454pp; English.
XX
XX The invention relates to a number of isolated Arabidopsis thaliana cDNA
CC sequences and their encoded proteins which are especially transcription
CC factor related cDNA's and proteins. The isolated or recombinant plant
CC transcription factor polynucleotides and polypeptides are useful in
CC producing transgenic plants with commercially valuable properties, i.e.
CC modified or altered desirable traits as compared to a reference plant,
CC such as an alteration in a plant growth characteristic, e.g. growth rate,
CC germination rate of seeds, vigor of plants and seedlings, or leaf and
CC flower senescence. Sequence information related to the polynucleotides
CC and polypeptides can also be used in bioinformatic search methods. The
CC transgenic plant is useful for growing a progeny plant from a parent
CC plant. This sequence represents one of the cDNAs of the invention.
XX
SQ Sequence 1304 BP; 391 A; 322 C; 238 G; 353 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 115 Length: 1304
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 63.64% Indels: 0
DB: 10 Gaps: 0

US-09-851-138C-155 (1-11) x ADD31048 (1-1304)

Qy 5 GlyAspIleIleLeuHisLeu 11
|||||
Db 1067 GGAGACATTATTCCTCATCTT 1047

RESULT 14
ADE31460/C
ID ADE31460 standard; cDNA; 1304 BP.
XX
AC ADE31460;
XX
DT 29-JAN-2004 (first entry)

XX Plant yield related polynucleotide clone G1052.
 DE ds; gene; transcription factor; transgenic plant; salt stress resistance;
 XX osmotic stress resistance; freezing tolerance; drought tolerance;
 KW low humidity tolerance; radiation resistance.
 KW Arabidopsis thaliana.
 OS
 FH Key Location/Qualifiers
 FT CDS 138..1127
 FT /*tag= a
 FT /product= "transcription factor"
 PN WO2003013228-A2.
 XX
 XX 20-FEB-2003.
 XX
 XX 09-AUG-2002; 2002WO-US025808.
 XX
 XX 09-AUG-2001; 2001US-0310847P.
 PR 19-NOV-2001; 2001US-0336049P.
 PR 11-DEC-2001; 2001US-0338692P.
 PR 14-JUN-2002; 2002US-00171468.
 XX
 XX (MEND-) MENDEL BIOTECHNOLOGY INC.
 PA
 XX Heard JE, Riechmann JL, Creelman RA, Keddie J, Pilgrim ML;
 PI Dubell AN, Jiang C, Ratcliffe O, Pineda O, Yu G, Broun PE;
 XX
 DR WPI; 2003-248222/24.
 DR P-PSDB; ADE31461.
 XX
 XX New plant transcription factor polynucleotides and polypeptides, useful
 PT in producing transgenic plants with commercially valuable properties,
 PT i.e. modified desirable traits, e.g. salt stress resistance or tolerance
 PT to freezing.
 XX
 PS Disclosure; SEQ ID NO 27; 311pp; English.
 XX
 XX The invention relates to a number of isolated cDNA sequences and their
 CC encoded proteins which are especially transcription factor related cDNA's
 CC and proteins. The isolated or recombinant plant transcription factor
 CC polynucleotides and polypeptides are useful in producing transgenic
 CC plants with commercially valuable properties, i.e. modified or altered
 CC desirable traits as compared to a reference plant, e.g. salt stress
 CC resistance, osmotic stress resistance, tolerance to freezing, drought,
 CC low humidity tolerance, or radiation resistance. Sequence information
 CC related to the polynucleotides and polypeptides can also be used in
 CC bioinformatic search methods. The transgenic plant is useful for growing
 CC a progeny plant from a parent plant. This sequence represents one of the
 CC cDNAs of the invention
 XX
 SQ Sequence 1304 BP; 391 A; 322 C; 238 G; 353 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 115 Length: 1304
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 10 Gaps: 0

US-09-851-138C-155 (1-11) x ADE31460 (1-1304)

Qy 5 GlyAspIleIleuIstleu 11
 |||||
 Db 1067 GGAGACATTATTCCTCATCTT 1047
 |||||
 RESULT 15
 ADI41750/c
 ID ADI41750 standard; DNA; 1304 BP.
 XX

AC ADI41750;
 XX
 DT 22-APR-2004 (first entry)
 XX
 DE Plant transcription factor polynucleotide #107.
 XX
 KW transgenic; plant; enhanced tolerance to abiotic stress;
 KW glycosylase tolerance; hormone sensitivity; disease resistance;
 KW sugar sensing; flowering; flower structure; stem bifurcation;
 KW branching pattern; apical dominance; trichome; stem morphology;
 KW root growth; root hair; seed development; cell proliferation;
 KW cell differentiation; premature senescence; necrosis; plant size;
 KW leaf morphology; seed morphology; seed biochemistry; root anthocyanin;
 KW plant anthocyanin; light response; shade avoidance; bioinformatic;
 KW transcription factor; gene; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN US2004019927-A1.
 XX
 XX 29-JAN-2004.
 XX
 XX 25-FEB-2003; 2003US-00374780.
 XX
 XX 18-APR-2001; 2001US-00837944.
 XX
 XX (SHER) SHERMAN B K.
 PA (RIEC) RIECHMANN J L.
 PA (JIAN) JIANG C.
 PA (HEAR) HEARD J E.
 PA (HAAR) HAAKE V.
 PA (CREE) CREELMAN R A.
 PA (RATC) RATCLIFFE O.
 PA (ADAM) ADAM L J.
 PA (REUB) REUBER T L.
 PA (KEDD) KEDDIE J.
 PA (BROU) BROUN P E.
 PA (PILG) PILGRIM M L.
 PA (DUBE) DUBELL A N.
 PA (PINE) PINEDA O.
 PA (YUGG) YU G.
 XX
 XX Sherman BK, Riechmann JL, Jiang C, Heard JE, Haake V;
 PI Creelman RA, Ratcliffe O, Adam LJ, Reuber TL, Keddie J, Broun PE;
 PI Pilgrim ML, Dubell AN, Pineda O, Yu G;
 XX
 DR WPI: 2004-132245/13.
 DR P-PSDB; ADI41751.
 XX
 XX New transgenic plant comprising a recombinant polynucleotide of any one
 PT of more than 500 nucleotide sequences, useful in bioinformatic search
 PT methods.
 XX
 PS Claim 1; SEQ ID NO 213; 435pp; English.
 XX
 XX The invention describes a transgenic plant comprising a recombinant
 CC polynucleotide of any one of more than 500 nucleotide sequences fully
 CC defined in the specification or its complement. The method of the
 CC invention can be used to produce a plant having altered traits such as:
 CC enhanced tolerance to abiotic stress; glycosylase tolerance; hormone
 CC sensitivity; disease resistance; sugar sensing; early or late flowering;
 CC altered flower structure, change in stem bifurcations, altered branching
 CC pattern, reduced apical dominance, reduced trichome density; lack of
 CC trichomes; reduced ectopic trichome development; altered trichome
 CC development; increase in trichome number; altered stem morphology;
 CC increased root growth; increased root hairs; altered seed development;
 CC altered cell proliferation or cell differentiation; rapid development;
 CC premature senescence; increased necrosis; increase in seedling or plant
 CC size; decreased plant size; leaf morphology; seed morphology; seed
 CC biochemistry; increase in root anthocyanins; increase in plant
 CC anthocyanins; or alteration in light response or shade avoidance. The
 CC transgenic plant, polynucleotides and polypeptides are useful in
 CC bioinformatic search methods. This sequence encodes a plant transcription

CC factor that can be used in the creation of a transgenic plant with
CC altered traits.

XX
SQ Sequence 1304 BP; 391 A; 322 C; 238 G; 353 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	115	Length:	1304
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	63.64%	Indels:	0
DB:	12	Gaps:	0

US-09-851-138C-155 (1-11) x ADI41750 (1-1304)

Oy 5 GlyAspIleIleLeuHisLeu 11

Db 1067 GGAGACATTATTCCTCATCTT 1047

Search completed: March 3, 2005, 16:26:08
Job time : 81.4667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:54:32 ; Search time 22.6769 Seconds
(without alignments)
793.716 Million cell updates/sec

Title: US-09-851-138c-155
Perfect score: 11
Sequence: 1 VYAGDIILHL 11

Scoring table: QXIGO,
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2395798

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlp
-MODEL=frame+ p2n model
-Q/cgn2_1/USPTO_spool_p/US09851138/runat_28022005_120307_21495/app_query.fasta_1.1123
-DB=issued Patents NA -Qfmt=fastap -SUFFIX=olog.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09851138 @CGN 1.1 249 @runat_28022005_120307_21495 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELOP=6 -DELEXT=7

Database : Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCUTS COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	100.0	447	US-08-836-075A-51	Sequence 51, Appl
2	7	63.6	2255	US-09-807-757C-10	Sequence 10, Appl
3	7	63.6	24338	US-09-949-016-12087	Sequence 12087, A
4	7	63.6	24639	US-09-949-016-15749	Sequence 15749, A
5	6	54.5	60	US-09-023-228B-126	Sequence 126, App
6	54.5	60	3	US-09-163-025B-126	Sequence 126, App
7	6	54.5	60	US-10-037-282-126	Sequence 126, App
8	6	54.5	62	US-08-697-631A-17	Sequence 17, Appl
9	6	54.5	210	US-09-583-110-105	Sequence 105, Appl
10	6	54.5	213	US-09-107-433-62	Sequence 62, Appl
11	6	54.5	350	US-09-513-999C-1410	Sequence 1410, App
12	6	54.5	427	US-09-060-756-191	Sequence 191, App

C 13	6	54.5	427	4	US-09-670-314-191	Sequence 191, App
C 14	6	54.5	601	4	US-09-949-016-36270	Sequence 36270, A
C 15	6	54.5	601	4	US-09-949-016-77644	Sequence 77644, A
C 16	6	54.5	601	4	US-09-949-016-79089	Sequence 79089, A
C 17	6	54.5	601	4	US-09-949-016-90003	Sequence 90003, A
C 18	6	54.5	601	4	US-09-949-016-121963	Sequence 121963, A
19	6	54.5	601	4	US-09-949-016-135813	Sequence 135813, A
20	6	54.5	601	4	US-09-949-016-135930	Sequence 135930, A
21	6	54.5	601	4	US-09-949-016-136047	Sequence 136047, A
22	6	54.5	601	4	US-09-949-016-138527	Sequence 138527, A
23	6	54.5	601	4	US-09-949-016-138528	Sequence 138528, A
24	6	54.5	601	4	US-09-949-016-138529	Sequence 138529, A
25	6	54.5	601	4	US-09-949-016-138530	Sequence 138530, A
26	6	54.5	601	4	US-09-949-016-145643	Sequence 145643, A
27	6	54.5	601	4	US-09-949-016-145644	Sequence 145644, A
28	6	54.5	601	4	US-09-949-016-145645	Sequence 145645, A
29	6	54.5	601	4	US-09-949-016-145911	Sequence 145911, A
30	6	54.5	601	4	US-09-949-016-145912	Sequence 145912, A
31	6	54.5	601	4	US-09-949-016-145913	Sequence 145913, A
32	6	54.5	601	4	US-09-949-016-146179	Sequence 146179, A
33	6	54.5	601	4	US-09-949-016-146180	Sequence 146180, A
34	6	54.5	601	4	US-09-949-016-146181	Sequence 146181, A
35	6	54.5	601	4	US-09-949-016-149794	Sequence 149794, A
36	6	54.5	601	4	US-09-949-016-159154	Sequence 159154, A
37	6	54.5	601	4	US-09-949-016-163072	Sequence 163072, A
38	6	54.5	601	4	US-09-949-016-183819	Sequence 183819, A
39	6	54.5	601	4	US-09-949-016-189592	Sequence 189592, A
40	6	54.5	601	4	US-09-949-016-198485	Sequence 198485, A
41	6	54.5	601	4	US-09-949-016-198486	Sequence 198486, A
42	6	54.5	601	4	US-09-949-016-203896	Sequence 203896, A
43	6	54.5	676	4	US-09-976-594-823	Sequence 823, App
44	6	54.5	825	4	US-09-489-039A-5829	Sequence 5829, App
45	6	54.5	921	4	US-09-543-681A-134	Sequence 134, App

ALIGNMENTS

RESULT 1
US-08-836-075A-51
; Sequence 51, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAETIENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; TITLE OF INVENTION: AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESS: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:

```
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-836-075A-51

Alignment Scores:
Pred. No.: 0.000428 Length: 447
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-851-138C-155 (1-11) x US-08-836-075A-51 (1-447)

QY 1 ValTyrGluAlaGlyAspIleIleLeuHisLeu 11
Db 160 GTGTATGAGCGCGGATATTATCTCCACTTA 192

RESULT 2
US-09-807-757C-10
; Sequence 10, Application US/09807757C
; Patent No. 6825035
; GENERAL INFORMATION:
; APPLICANT: Owens, Gary K.
; APPLICANT: Mack, Christopher
; APPLICANT: Blank, Randall
; APPLICANT: University of Virginia Patent Foundation
; TITLE OF INVENTION: Compositions and Methods for Modulating Expression
; FILE REFERENCE: 021258-00050005
; CURRENT APPLICATION NUMBER: US/09/807,757C
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 60/105,330
; PRIOR FILING DATE: 1998-10-23
; PRIOR APPLICATION NUMBER: WO PCT/US99/24972
; PRIOR FILING DATE: 1999-10-22
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 2255
; TYPE: DNA
; ORGANISM: Gallus sp.
; FEATURE:
; OTHER INFORMATION: smooth muscle alpha-actin (SM alpha-A) gene
; OTHER INFORMATION: first intron sequence
US-09-807-757C-10

Alignment Scores:
Pred. No.: 52.4 Length: 2255
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 63.64% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-155 (1-11) x US-09-807-757C-10 (1-2255)

QY 3 GluAlaGlyAspIleIleLeu 9
Db 1923 GAGGCGAGGGGACATCATCTG 1943

RESULT 3
US-09-949-016-12087
; Sequence 12087, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12087
; LENGTH: 24638
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12087

Alignment Scores:
Pred. No.: 460 Length: 24638
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 63.64% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-155 (1-11) x US-09-949-016-12087 (1-24638)

QY 4 AlaGlyAspIleIleLeuHis 10
Db 4298 GCGGGGACATTATCTGCAT 4318

RESULT 4
US-09-949-016-15749
; Sequence 15749, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15749
; LENGTH: 24639
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15749

Alignment Scores:
Pred. No.: 460 Length: 24639
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 63.64% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-155 (1-11) x US-09-949-016-15749 (1-24639)

QY 4 AlaGlyAspIleIleLeuHis 10
```

Db 4298 GCGGGGACATTCTGTCAT 4318
RESULT 5
US-09-023-228B-126/c
; Sequence 126, Application US/09023228B
; Patent No. 6140490
; GENERAL INFORMATION:
; APPLICANT: BIESECKER, GREGORY
; TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF
; TITLE OF INVENTION: COMPLEMENT SYSTEM PROTEINS
; NUMBER OF SEQUENCES: 157
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson & Bratschun, L.L.C.
; STREET: 8400 E. Prentice Place #200
; CITY: Denver
; STATE: Colorado
; COUNTRY: US
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,228B
; FILING DATE: 12-FEBRUARY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/01739
; FILING DATE: 30 JAN 1997
; PRIOR APPLICATION NUMBER: 08/595,335
; FILING DATE: 1 FEB 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson, Esq.
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX50/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 126:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA
; FEATURE:
; OTHER INFORMATION: All pyrimidines are 2'-F modified
US-09-023-228B-126
Alignment Scores:
Pred. No.: 25.2 Length: 60
Score: 6.00 Matches: 6
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Indels: 0
DB: 0 Gaps: 0
US-09-851-138C-155 (1-11) x US-09-023-228B-126 (1-60)
QY 1 ValTYrGluAlaGlyAsp 6
Db 49 GTCTACGAGCTGCTGAC 32
RESULT 6
US-09-163-025B-126/c
; Sequence 126, Application US/09163025B
; Patent No. 6395888
; GENERAL INFORMATION:
; APPLICANT: Nexstar Pharmaceuticals, Inc.
; TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF COMPLEMENT SYSTEM

; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: NEX 50 CIP2
; CURRENT APPLICATION NUMBER: US/09/163,025B
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 09/023,228
; PRIOR FILING DATE: 1998-02-12
; PRIOR APPLICATION NUMBER: PCT/US97/01739
; PRIOR FILING DATE: 1997-01-30
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 126
; LENGTH: 60
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Completely
; NAME/KEY: modified_base
; LOCATION: (1)..(60)
; OTHER INFORMATION: All c's and u's are 2'-F
US-09-163-025B-126
Alignment Scores:
Pred. No.: 25.2 Length: 60
Score: 6.00 Matches: 6
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Indels: 0
DB: 0 Gaps: 0
US-09-851-138C-155 (1-11) x US-09-163-025B-126 (1-60)
QY 1 ValTYrGluAlaGlyAsp 6
Db 49 GTCTACGAGCTGCTGAC 32
RESULT 7
US-10-037-282-126/c
; Sequence 126, Application US/10037282
; Patent No. 6566343
; GENERAL INFORMATION:
; APPLICANT: BIESECKER, GREGORY
; APPLICANT: GOLD, LARRY
; TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF COMPLEMENT SYSTEM
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: NEX 50 CIP 2 CON
; CURRENT APPLICATION NUMBER: US/10/037,282
; CURRENT FILING DATE: 2002-01-03
; PRIOR APPLICATION NUMBER: 09/163,025
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 08/595,335
; PRIOR FILING DATE: 1996-02-01
; PRIOR APPLICATION NUMBER: PCT/US97/01739
; PRIOR FILING DATE: 1997-01-30
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 126
; LENGTH: 60
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Nucleic Acid
; OTHER INFORMATION: all pyrimidines are 2'F
US-10-037-282-126
Alignment Scores:
Pred. No.: 25.2 Length: 60
Score: 6.00 Matches: 6
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Indels: 0
DB: 4 Gaps: 0

APPLICANT:	Gordon, Stephen
APPLICANT:	Billault, Alain
TITLE OF INVENTION:	METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM ANTI-SENSE: NO
TITLE OF INVENTION:	THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
TITLE OF INVENTION:	LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
FILE REFERENCE:	3495-0169
CURRENT APPLICATION NUMBER:	US/09/060,756
CURRENT FILING DATE:	1998-04-16
NUMBER OF SEQ ID NOS:	743
SOFTWARE:	PatentIn Ver. 2.0
SEQ ID NO 191	
LENGTH:	427
TYPE:	DNA
ORGANISM:	Mycobacterium tuberculosis
FEATURE:	
NAME/KEY:	unsure
LOCATION:	(various positions within the sequence)
OTHER INFORMATION:	applicants are uncertain of bases designated as "n"
US-09-060-756-191	
Alignment Scores:	
Pred. No.:	150
Score:	6.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	54.55%
DB:	3
US-09-851-138C-155 (1-11) x US-09-060-756-191 (1-427)	
QY	1 ValTyrGluAlaGlyAsp 6
Db	27 GTTATGAGCAGGTGAT 10
RESULT 13	
US-09-670-314-191/c	
Sequence 191, Application US/09670314	
Patent No. 6492506	
GENERAL INFORMATION:	
APPLICANT: Cole, Stewart	
APPLICANT: Buchrieser-Brosch, Roland	
APPLICANT: Gordon, Stephen	
APPLICANT: Billault, Alain	
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM	
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA	
TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA	
FILE REFERENCE: 3495-0169	
CURRENT APPLICATION NUMBER: US/09/670,314	
CURRENT FILING DATE: 2001-01-12	
PRIOR APPLICATION NUMBER: 09/060,756	
PRIOR FILING DATE: 1998-04-16	
NUMBER OF SEQ ID NOS: 743	
SOFTWARE: PatentIn Ver. 2.0	
SEQ ID NO 191	
LENGTH: 427	
TYPE: DNA	
ORGANISM: Mycobacterium tuberculosis	
FEATURE:	
NAME/KEY: unsure	
LOCATION: (various positions within the sequence)	
OTHER INFORMATION: applicants are uncertain of bases designated as "n"	
US-09-670-314-191	
Alignment Scores:	
Pred. No.:	150
Score:	6.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	54.55%
DB:	4
US-09-851-138C-155 (1-11) x US-09-670-314-191 (1-427)	

Qy		Query Match:	54.55%	Indels:	0
		DB:	4	Gaps:	0
1	VAltyrGluAlaGLyAsp 6				
27	GTTTATGAAGCAGGTGAT 10	US-09-851-138C-155 (1-11) x US-09-949-016-77644 (1-601)			

		Qy	6 AspIleLeuHisLeu 11
RESULT 14			
US-09-949-016-36270/c			
: Sequence 36270, Application US/09949016			
: Parent No C812139		Dd	478 GATATACTTACACTA 461

APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE. METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: C20001197
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36270
; LENGTH: 601
;

```

US-09-949-016-36270

Alignment Scores:	
Pred. No.:	204
Length:	601
Match:	500
Mismatch:	6
Gap:	6

Conservative:	0
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Mismatches:	0
Indels:	0
Query Match:	54.55%
Gap:	4
Gap:	0

US-09-851-138C-155 (1-11) x US-09-949-016-36270 (1-601)
Ov 6 AsaT1e1LeuH1a1eU 11

Db
40 GATATAATTCTACACCTA 23

RESULT 15
US-09-949-016-77644/c
; Sequence 77644, Application US/09949016

```

: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: WITH TRACED METHODS OF DETECTION AND USES THEREOF

```

FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIORITY APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03

FOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012

```

, SEQ ID NO 77644
,
, LENGTH: 601
,
, TYPE: DNA
,

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:43:48 ; Search time 693.169 Seconds
(without alignments)
604.047 Million cell updates/sec

Title: US-09-851-138c-155
Perfect score: 11
Sequence: 1 VYAGDIHL 11

Scoring table: OLIGO³
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Word size: 1

Total number of hits satisfying chosen parameters: 68471649

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n model -DRV=xlp
-Q/cn2_1/USPTO_spool_p/US09851138/runat_28022005_120306_21476/app_query.fasta_1.1123
-DB=EST -QFMT=fastap -SUFFIX=olig.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09851138 @CCN 1.1 10973 @runat_28022005_120306_21476 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -YGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	8	72.7	275	1 AA249854	AA249854 m0989.seq
C 2	8	72.7	334	7 CO224063	CO224063 WS01020.B
C 3	7	63.6	191	7 H55311	H55311 CHR220250 C
C 4	7	63.6	341	9 CG509249	CG509249 OST60251
C 5	7	63.6	376	5 BY038664	BY038664 BY038664
C 6	7	63.6	396	5 BP669753	BP669753 BP669753
C 7	7	63.6	402	1 AV801673	AV801673 AV801673
C 8	7	63.6	403	8 BH019452	BH019452 L3624b.d
C 9	7	63.6	408	5 BP594076	BP594076 BP594076

10	7	63.6	410	5	BP672219	BP672219
11	7	63.6	415	5	BP670021	BP670021
12	7	63.6	421	5	BP648445	BP648445
13	7	63.6	422	1	AV820556	AV820556
14	7	63.6	424	5	BP597828	BP597828
15	7	63.6	425	8	BH019424	BH019424
16	7	63.6	430	5	BP603275	BP603275
C 17	7	63.6	433	8	AQ438695	AQ438695
C 18	7	63.6	434	8	BH019453	BH019453
C 19	7	63.6	451	1	AI704112	AI704112
C 20	7	63.6	453	8	B88842	B88842
C 21	7	63.6	459	5	BP642414	BP642414
C 22	7	63.6	473	8	BH018847	BH018847
C 23	7	63.6	491	8	CM485047	CM485047
C 24	7	63.6	491	8	BA7990	BA7990
C 25	7	63.6	492	8	AQ661807	AQ661807
C 26	7	63.6	519	7	CO794439	CO794439
C 27	7	63.6	523	8	AZ699957	AZ699957
C 28	7	63.6	525	8	AQ808205	AQ808205
C 29	7	63.6	542	9	CE184041	CE184041
C 30	7	63.6	556	2	BE187690	BE187690
C 31	7	63.6	567	8	AZ154671	AZ154671
C 32	7	63.6	578	6	CA991526	CA991526
C 33	7	63.6	579	1	AI727146	AI727146
C 34	7	63.6	584	7	COL32821	COL32821
C 35	7	63.6	586	1	AV439683	AV439683
C 36	7	63.6	588	6	CA991525	CA991525
C 37	7	63.6	592	8	AZ321487	AZ321487
C 38	7	63.6	592	8	BH019349	BH019349
C 39	7	63.6	592	9	CE346616	CE346616
C 40	7	63.6	600	8	AQ158531	AQ158531
C 41	7	63.6	603	2	AW218146	AW218146
C 42	7	63.6	605	2	BE432913	BE432913
C 43	7	63.6	615	9	CE414425	CE414425
C 44	7	63.6	617	1	AI730649	AI730649
C 45	7	63.6	619	9	CR188957	CR188957

ALIGNMENTS

RESULT 1	AA249854	275 bp	mRNA	linear	EST 11-MAR-1997
AA249854/c	m0989.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA				
LOCUS	5', mRNA sequence.				
DEFINITION	AA249854.1 GI:1881141				
ACCESSION	AA249854				
VERSION	AA249854.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 275)				
AUTHORS	Liew, C.C.				
TITLE	cDNAs from human fetal heart (1997)				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Liew CC Brigham and Women's Hospital Harvard Medical School 75 Francis St. Boston, MA 02115, USA Tel: 617/7328915 Fax: 617/9750995 Email: cliew@rics.bwh.harvard.edu				

PCR Primers
FORWARD: 5' GCCAGCTCGAAATTAACCTCCTACTAAAGG 3'
BACKWARD: 5' CCAGTGAATTGTAATACGACTCCTATAGGGC 3'
Seq primer: 5' GAAATTAACCTCCTACTAAAGG 3'.
Location/Qualifiers
1..275
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/lab_host="E. coli XLI-Blue"

/clone_lib="Human fetal heart, Lambda ZAP Express"
 /note="Vector: Lambda ZAP Express; Site 1: EcoRI; Site 2:
 XhoI; mRNA was purified from human fetal hearts (8-10
 weeks). cDNA was synthesized using a XhoI-Oligo dT
 adaptor-primer. EcoRI adaptors were ligated, followed by
 digestion with XhoI, for directional cloning into
 predigested lambda ZAP Express."

ORIGIN

Alignment Scores:
 Pred. No.: 33.3 Length: 275
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 72.73% Indels: 0
 DB: 1 Gaps: 0

US-09-851-138C-155 (1-11) x AA249854 (1-275)

Qy 4 AlaGlyAspIleLeuHisLeu 11
 |||||
 Db 211 GCGGAGATATATCTTACATT 188

RESULT 2

CO224063/c
 LOCUS CO224063 334 bp mRNA linear EST 22-JUN-2004
 DEFINITION WS01020.B21 L18 SS-R-N-A-11 Picea sitchensis cDNA clone WS01020_L18
 3', mRNA sequence.

ACCESSION CO224063

VERSION CO224063.1 GI:49046378

KEYWORDS EST.

SOURCE Picea sitchensis (Sitka spruce)

ORGANISM Picea sitchensis

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatothya; Coniferopsida; Coniferales; Pinaceae; Picea.

AUTHORS

1 (bases 1 to 334)
 Ralph, S., Kolosova, N., Cooper, D., Butterfield, Y., Kirkpatrick, R.,
 Liu, J., Palquist, D., Stott, J., Barber, S., Yang, G., Babakalif, R.,
 Brown-John, M., Chand, S., Featherstone, R., Masson, A., Mayo, M.,
 Moran, J., Olson, T., Wong, D., Friedmann, M.F., Ritland, C.E.,
 Siddiqui, A., Holt, R., Jones, S., Marra, M., Ellis, B.E., Douglas, C.,
 Ritland, K. and Bohlmann, J.

TITLE The spruce transcriptome: Analysis of expressed sequence tags from
 multiple cDNA libraries

JOURNAL

Unpublished (2004)

COMMENT

Contact: Joerg Bohlmann
 Genome BC forest genomics program
 University of British Columbia
 UBC Biotechnology Laboratory, 6174 University Boulevard, Rm. 237,
 Vancouver, British Columbia, Canada, V6T 1Z3
 Tel: 1-604-822-0282
 Fax: 1-604-822-6097
 Email: bohlmann@interchange.ubc.ca
 Plate: WS01020 row: L column: 18
 High quality sequence stop: 334
 POLYA=Yes.

FEATURES

source

1..334 Location/Qualifiers

/organism="Picea sitchensis"

/mol_type="mRNA"

/cultivar="GB2-229"

/db_xref="taxon:3332"

/clone="WS01020 L18"

/sex="Hermaprodite"

/tissue_type="Young root growth (terminal 1-3 cm) and old

root growth (discal to terminal 1-3 cm) tissues"

/dev_stage="three year old clonal trees grown under

greenhouse conditions in standard potting soil mixture."

/lab_host="E. coli DH10B cells"

/clone_lib="SS-R-N-A-11"

/note="Organ: Roots; Vector: pBluescript II SK (+) XR;

Site1: EcoRI (5' end of cDNA); Site2: XhoI (3' end of

cDNA); mRNA was isolated from each tissue source

independently and equal quantities of mRNA from each
 tissue were then pooled. cDNA was prepared from 5
 micrograms of mRNA and directionally ligated into the
 pBluescript II SK (+) XR vector using the pBluescript II
 X8 cDNA Library Construction Kit according to
 manufacturer's instructions with modifications
 (Stratagene). Plasmid DNA was then transformed by
 electroporation into DH10B cells (Invitrogen) for
 propagation. Normalization was applied according to
 published methods [Bonaldo M.F. et al. (1996) Genome
 Research 6(9):791] in order to reduce the abundance of
 highly expressed transcripts."

ORIGIN

Alignment Scores:
 Pred. No.: 39 Length: 334
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 72.73% Indels: 0
 DB: 7 Gaps: 0

US-09-851-138C-155 (1-11) x CO224063 (1-334)

Qy 4 AlaGlyAspIleLeuHisLeu 11
 |||||
 Db 203 GCGGAGATATATCTTACATT 180

RESULT 3

H55311

LOCUS

DEFINITION H55311 191 bp mRNA linear EST 07-DEC-1995
 CHR220250 Chromosome 22 exon Homo sapiens cDNA clone C22_310 5',
 mRNA sequence.

ACCESSION H55311

VERSION H55311.1 GI:1108177

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 191)

Trofatter, J.A., Long, K.R., Murrell, J.R., Stotler, C.J., Gusella, J.F.
 and Buckler, A.J.

TITLE An expression-independent catalog of genes from human chromosome 22

JOURNAL Genome Res. 5 (3), 214-224 (1995)

MEDLINE 96159527

PUBMED 8593609

COMMENT

Contact: Buckler AJ
 Molecular Neurogenetics Unit
 Massachusetts General Hospital
 Building 149, 13th St., Charlestown MA 02129
 Tel: 6177249616
 Fax: 6177265736
 Email: buckler@helix.mgh.harvard.edu
 Seq primer: T3.

FEATURES

source

1..191 Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="C22_310"

/lab_host="E. coli DH5a"

/clone_lib="Chromosome 22 exon"

/note="Vector: pBluescriptIIKS+; Site 1: Sal I; Site 2:

Bam HI (destroyed); Exons were isolated from human

Chromosome 22 specific cosmids using a modification of

the method of exon amplification (Proc. Natl. Acad. Sci.

USA 88:4005-4009, 1991). Amplified exons were digested

with Sal I and Bgl II and subsequently cloned into

pBluescriptIIKS+ at the Sal I and Bam HI sites."

ORIGIN

Alignment Scores:

Pred. No.: 288 Length: 191
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 7 Gaps: 0

US-09-851-138C-155 (1-11) x H55311 (1-191)

QY 5 GlyAspIleLeuHisLeu 11

DB 44 GTGACATAATCTTCATCTT 64

RESULT 4

CG509249/c

LOCUS

DEFINITION CG509249 341 bp mRNA linear GSS 01-OCT-2003

OST60251 Mus musculus 129Sv/Ev Mus musculus cDNA clone OST60251,

mRNA sequence.

CG509249

VERSION GSS.

KEYWORDS CG509249.1 GI:37292297

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 341)

REFERENCE Zambrowicz,B.P., Abuin,A., Ramirez-Solis,R., Richter,L.J.,

Piggott,J., Beltranda,Rio,H., Buxton,E.C., Edwards,J., Finch,R.A.,

Fridde,C.J., Gupta,A., Hansen,G., Hu,Y., Huang,W., Jaing,C.,

Key,B.W. Jr., Kipp,P., Kohlhaufl,B., Ma,Z.-Q., Markesich,D.,

Payne,R., Potter,D.G., Qian,N., Shaw,J., Schrick,J., Shi,Z.-Z.,

Sparks,M.J., Van Sligtenhorst,I., Vogel,P., Walke,W., Xu,N.,

Zhu,Q., Person,C. and Sands,A.T.

Wnk1 kinase deficiency lowers blood pressure in mice: a gene-trap

screen to identify potential targets for therapeutic intervention

Proc. Natl. Acad. Sci. U.S.A. 100 (24), 14109-14114 (2003)

CONTACT: Zambrowicz BP

Omnibank

Lexicon Genetics Incorporated

4000 Research Forest Drive, The Woodlands, TX 77381, USA

Email: materials@lexgen.com

Gene trap sequence tag generated by 3' RACE from mouse ES cells as

described in Zambrowicz et al (Nature. 1998 Apr 9;392(6676): 608-11)

Class: Gene Trap.

Location/Qualifiers

1. 341

/organism="Mus musculus"

/mol_type="mRNA"

/strain="129Sv/Ev"

/db_xref="taxon:10090"

/clone="OST60251"

/cell_type="embryonic stem cell"

/clone_lib="Mus musculus 129Sv/Ev"

ORIGIN

Alignment Scores:

Pred. No.: 461 Length: 341

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 63.64% Indels: 0

DB: 9 Gaps: 0

US-09-851-138C-155 (1-11) x CG509249 (1-341)

QY 1 ValTyrGluAlaGlyAspIle 7

DB 196 GTGTATGAAGCTGGAGACATC 176

RESULT 5

BY038664

LOCUS

DEFINITION BY038664 RIKEN full-length enriched, pooled tissues,

376 bp mRNA linear EST 06-DEC-2002

BY038664

BY038664 RIKEN full-length enriched, pooled tissues,

376 bp mRNA linear EST 06-DEC-2002

cell_line=TIB-55BB88, etc. Mus musculus cDNA clone I730007B04 5',
 mRNA sequence.

BY038664

ACCESSION BY038664.1 GI:26144107

VERSION

KEYWORDS

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 376)

REFERENCE

AUTHORS

Okazaki,Y., Furuno,M., Saito,R., Suzuki,H., Yamanaka,I.,

Nikaido,I., Oeato,N., Kato,T., Tomaru,Y., Hasegawa,Y., Nogami,A.,

Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,

Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,

Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,

Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,

Chavakis,C., Corbelli,L.E., Cousins,S., Dalla,E., Dragani,T.A.,

Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,

Gariboldi,M., Giasi,C., Godzik,A., Gough,J., Grimond,S.,

Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,

Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,

Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,

Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,

Nunata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,

Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,

Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,

Sandelin,A., Schneider,C., Sempke,C.A., Setou,M., Shinada,K.,

Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R., Tomita,M.,

Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,

Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,

Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,

Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,

Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,

Arakawa,T., Fukuda,S., Hara,A., Hashizume,M., Imotani,K., Ishii,Y.,

Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,

Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,

Rogers,J., Birney,E. and Hayashizaki,Y.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

CONTACT: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/

Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,

Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,

Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,

Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,

Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and

Hayashizaki,Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multipipillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in Riken.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

FEATURES
source

```

1. .376
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, pooled tissues,
cell_line=TIB-55BB88, etc."
/notes="pooled tissues: (sex=mix, cell_line=TIB-55BB88),
(sex=mix, cell_line=CRJ-1722 L5178Y-R)"

```

ORIGIN

Alignment Scores:
 Pred. No.: 499 Length: 376
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 5 Gaps: 0

US-09-851-138C-155 (1-11) x BY038664 (1-376)

Qy 5 GlyAspIleIleLeuHisLeu 11
 |||||
 Db 290 GGTGACATTATCTCCACCTC 310

RESULT 6

BP669753
 LOCUS BP669753 396 bp mRNA linear EST 28-JUN-2004
 DEFINITION BP669753 RAFL21 Arabidopsis thaliana cDNA clone RAFL21-32-J10 3',
 mRNA sequence.

ACCESSION BP669753.1 GI:49321256
 VERSION BP669753
 KEYWORDS EST.

SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 396)
 AUTHORS Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
 Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
 Hayaehizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
 Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.

TITLE Functional annotation of a full-length Arabidopsis cDNA collection

JOURNAL Science 296 (5565), 141-145 (2002)
 MEDLINE 21932900
 PUBMED 11910074

COMMENT Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: mseki@rtc.riken.go.jp
 reversed clone; please visit our web site
 (http://pfweb.gsc.riken.go.jp/) for further details.

FEATURES
source

```

1. .396
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL21-32-J10"
/lab_host="DH10B"
/clone_lib="RAFL21"
/notes="Site 1: BamHI; Site 2: SalI; Subtraction library.  

The sequence was obtained from samples subjected to  

various stress and plant hormones-treated"

```

ORIGIN

Alignment Scores:
 Pred. No.: 520 Length: 396

Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 5 Gaps: 0

US-09-851-138C-155 (1-11) x BP669753 (1-396)

Qy 5 GlyAspIleIleLeuHisLeu 11
 |||||
 Db 319 GGAGACATTATCTCCATCTT 339

RESULT 7

AV801673
 LOCUS AV801673 402 bp mRNA linear EST 29-MAR-2002
 DEFINITION AV801673 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-29-C05 3',
 mRNA sequence.

ACCESSION AV801673
 VERSION AV801673.1 GI:19835658

KEYWORDS EST.

SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 402)

AUTHORS Seki,M., Narusaka,M., Iehida,J., Kamiya,A., Satou,M., Nakajima,M.,
 Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
 Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayaehizaki,Y.
 and Shinozaki,K.

TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)

JOURNAL Unpublished (2002)

COMMENT Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: mseki@rtc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially
 as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
 and XhoI was ligated to modified lambda FLC-1 vector (Carninci et
 al., submitted for publication) digested with BamHI and SalI. This
 clone is in a modified pBluescript vector. Please visit our web
 site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
 details.

FEATURES
Location/Qualifiers

```

1. .402
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL09-29-C05"
/dev_stage="plants at various developmental stages from  

germination to mature seeds"
/lab_host="DH10B"
/clone_lib="RAFL9"
/notes="Site 1: BamHI; Site 2: SalI; subjected to  

dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24  

hr) treatments"

```

ORIGIN

Alignment Scores:
 Pred. No.: 526 Length: 402
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 1 Gaps: 0

US-09-851-138C-155 (1-11) x AV801673 (1-402)

Qy 5 GlyAspIleIleLeuHisLeu 11
 |||||

```

Db      355 GGAGACATTATCTTCATCTT 375
RESULT 8
LOCUS   BH019452
DEFINITION
L3624b.d.HyGT3.2 Leishmania major Friedlin Cosmid Genomic Library
Leishmania major genomic clone L3624b, genomic survey sequence.
ACCESSION
VERSION BH019452
KEYWORDS GSS.
SOURCE   Leishmania major
ORGANISM Leishmania major
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE
1 (bases 1 to 403)
Miyler,P.J., Vogt,C., Cawthra,J., Klacking,M., Marty,A., Mack,J.,
Munden,H., Nguyen,D., Robertson,L., Sisk,E., Fazelinia,G.,
Aggarwal,G., Nelson,S., Seyler,A., Worthey,E. and Stuart,K.
Leishmania major Friedlin Cosmid End Sequences
Unpublished (2000)
Other GSSs: L3624b.d.HyGT7a.1
Contact: Myler PJ
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109-1651, USA
Tel: 206 284-8846
Fax: 206 284-0313
Email: mylerp@sbri.org
Seq primer: HyGT3
Class: cosmid ends.
FEATURES
source
Location/Qualifiers
1..403
/organism="Leishmania major"
/mol_type="genomic DNA"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="L3624b"
/lab_host="E. coli ED8767"
/clone_lib="Leishmania major Friedlin Cosmid Genomic
Library"
/note="Vector: cLHV3; Site 1: BamHI; Genomic DNA from
Leishmania major Friedlin was partially digested with
Sau3AI, size selected, and ligated with BamHI-digested
cLHV3 cosmid vector DNA. 9216 clones were picked and
arrayed. Library construction is described in Ivens et
al., Genomics Research, 8:135-145 (1998). The cLHV3
vector (Acc. No. CVU59231) is described in Ryan et al.,
Gene, 131:145-150 (1993)"
ORIGIN
Alignment Scores:
Pred. No.: 527 Length: 403
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 63.64% Indels: 0
DB: 8 Gaps: 0
US-09-851-138c-155 (1-11) x BH019452 (1-403)
QY 5 GlyAspIleIleLeuHisLeu 11
|||||
Db 234 GGGGACATTCATCTTCATCTC 254
RESULT 9
LOCUS   BP594076
DEFINITION
BP594076 RAPL15 Arabidopsis thaliana cDNA clone RAPL15-24-122 3',
mRNA sequence.
ACCESSION
VERSION BP594076.1 GI:49161544
KEYWORDS EST.
SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
1 (bases 1 to 410)
Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
Functional annotation of a full-length Arabidopsis cDNA collection
Science 296 (5565), 141-145 (2002)
BP594076 408 bp mRNA linear EST 23-JUN-2004
LOCUS   BP594076
DEFINITION
BP594076 RAPL15 Arabidopsis thaliana cDNA clone RAPL15-24-122 3',
mRNA sequence.
ACCESSION
VERSION BP594076.1 GI:49161544
KEYWORDS EST.
SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
1 (bases 1 to 410)
Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
Functional annotation of a full-length Arabidopsis cDNA collection
Science 296 (5565), 141-145 (2002)

```

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
1 (bases 1 to 408)
Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
Functional annotation of a full-length Arabidopsis cDNA collection
Science 296 (5565), 141-145 (2002)
21932900
MEDLINE
PUBMED
11910074
COMMENT
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Teikuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msekier@cc.riken.go.jp
reversed clone; please visit our web site
(http://pfigweb.gsc.riken.go.jp/) for further details.
FEATURES
source
Location/Qualifiers
1..408
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAPL15-24-122"
/tissue type="mixture of silique and flower"
/lab_host="DH10B"
/clone_lib="RAPL15"
/note="Site_1: BamHI; Site_2: SalI"
ORIGIN
Alignment Scores:
Pred. No.: 533 Length: 408
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 63.64% Indels: 0
DB: 5 Gaps: 0
US-09-851-138c-155 (1-11) x BP594076 (1-408)
QY 5 GlyAspIleIleLeuHisLeu 11
|||||
Db 356 GGAGACATTATCTTCATCTT 376
RESULT 10
LOCUS   BP672219
DEFINITION
BP672219 RAPL21 Arabidopsis thaliana cDNA clone RAPL21-43-G03 3',
mRNA sequence.
ACCESSION
VERSION BP672219.1 GI:49323722
KEYWORDS EST.
SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
1 (bases 1 to 410)
Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
Functional annotation of a full-length Arabidopsis cDNA collection
Science 296 (5565), 141-145 (2002)
21932900
MEDLINE
PUBMED
11910074
COMMENT
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Teikuba, Ibaraki 305-0074, Japan

```

Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: msek@rtc.riken.go.jp
 reversed clone; please visit our web site
 (http://pfweb.gsc.riken.go.jp/) for further details.

FEATURES

source

1. .410
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAFL21-43-G03"
 /lab_host="DH10B"
 /clone_lib="RAFL21"
 /note="Site 1: BamHI; Site 2: SalI; Subtraction Library.
 The sequence was obtained from samples subjected to
 various stress and plant hormones-treated"

ORIGIN

Alignment Scores:
 Pred. No.: 535 Length: 410
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 5 Gaps: 0

US-09-851-138C-155 (1-11) x BP672219 (1-410)

Qy 5 GlyAspIleIleLeuHisLeu 11

Db 338 GGAGACATTATTCCTTCATCTT 358

RESULT 11

BP670021

LOCUS BP670021 RAFL21 Arabidopsis thaliana cDNA clone RAFL21-33-H23 3',
 DEFINITION mRNA sequence.

VERSION BP670021.1 GI:49321524

KEYWORDS EST.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 415)

Authors: Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T.,
 Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M.,
 Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y.,
 Arakawa, T., Shibata, K., Shinagawa, A. and Shinozaki, K.

TITLE Functional annotation of a full-length Arabidopsis cDNA collection

JOURNAL Science 296 (5565), 141-145 (2002)

MEDLINE 21932900

PUBMED 11910074

COMMENT Contact: Motoaki Seki

Plant Functional Genomics Research Group

RIKEN Genomic Sciences Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-4359

Fax: 81-298-36-9060

Email: msek@rtc.riken.go.jp

reversed clone; please visit our web site

(http://pfweb.gsc.riken.go.jp/) for further details.

FEATURES

source

1. .415
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAFL21-33-H23"
 /lab_host="DH10B"
 /clone_lib="RAFL21"
 /note="Site 1: BamHI; Site 2: SalI; Subtraction Library.
 The sequence was obtained from samples subjected to

various stress and plant hormones-treated"

ORIGIN

Alignment Scores:
 Pred. No.: 540 Length: 415
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 5 Gaps: 0

US-09-851-138C-155 (1-11) x BP670021 (1-415)

Qy 5 GlyAspIleIleLeuHisLeu 11

Db 356 GGAGACATTATTCCTTCATCTT 376

RESULT 12

BP648445

LOCUS BP648445 RAFL19 Arabidopsis thaliana cDNA clone RAFL19-79-C08 3',
 DEFINITION mRNA sequence.

ACCESSION BP648445

VERSION BP648445.1 GI:49299915

KEYWORDS EST.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 421)

AUTHORS

Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T.,
 Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M.,
 Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y.,
 Arakawa, T., Shibata, K., Shinagawa, A. and Shinozaki, K.

TITLE Functional annotation of a full-length Arabidopsis cDNA collection

JOURNAL Science 296 (5565), 141-145 (2002)

MEDLINE 21932900

PUBMED 11910074

COMMENT Contact: Motoaki Seki

Plant Functional Genomics Research Group

RIKEN Genomic Sciences Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-4359

Fax: 81-298-36-9060

Email: msek@rtc.riken.go.jp

reversed clone; please visit our web site

(http://pfweb.gsc.riken.go.jp/) for further details.

FEATURES

source

1. .421
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAFL19-79-C08"
 /tissue_type="mixture of silique and flower"
 /lab_host="DH10B"
 /clone_lib="RAFL19"
 /note="Site 1: BamHI; Site 2: SalI; Subtraction Library"

ORIGIN

Alignment Scores:

Pred. No.: 546 Length: 421
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.64% Indels: 0
 DB: 5 Gaps: 0

US-09-851-138C-155 (1-11) x BP648445 (1-421)

Qy 5 GlyAspIleIleLeuHisLeu 11

Db 323 GGAGACATTATTCCTTCATCTT 343


```

RESULT 13
AV820556 422 bp mRNA linear EST 01-APR-2002
LOCUS AV820556 RAFL11 Arabidopsis thaliana cDNA clone RAFL11-11-A10 3',
DEFINITION mRNA sequence.
ACCESSION AV820556.1 GI:19862531
VERSION AV820556.1
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 422)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Sakurai,T.,
Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msek@tc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified lambda FLC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index\_e.html) for further
details.
FEATURES             Location/Qualifiers
     source           1..422
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL11-11-A10"
                     /lab_host="DH10B"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /clone_lib="RAFL11"
                     /note="Site 1: BamHI; Site 2: SalI; subjected to various
                     treatments (dehydration, cold, high salt, ABA, heat and
                     UV). Dark-grown plants"
ORIGIN
Alignment Scores: 547 Length: 422
Pred. No.: 7.00 Matches: 7
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 63.64% Gaps: 0
DB: 1
US-09-851-138C-155 (1-11) x AV820556 (1-422)
Qy 5 GlyAspIleLeuHisLeu 11
|||||
Db 324 GGAGACATTATTCTCCATCTT 344
|||||
RESULT 14
BP597828 424 bp mRNA linear EST 23-JUN-2004
LOCUS BP597828 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-01-C17 3',
DEFINITION mRNA sequence.
ACCESSION BP597828
VERSION BP597828.1 GI:49165296
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 422)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Sakurai,T.,
Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msek@tc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified lambda FLC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index\_e.html) for further
details.
FEATURES             Location/Qualifiers
     source           1..424
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL16-01-C17"
                     /lab_host="DH10B"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /clone_lib="RAFL16"
                     /note="Site 1: BamHI; Site 2: SalI; dark-grown"
ORIGIN
Alignment Scores: 550 Length: 424
Pred. No.: 7.00 Matches: 7
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 63.64% Gaps: 0
DB: 5
US-09-851-138C-155 (1-11) x BP597828 (1-424)
Qy 5 GlyAspIleLeuHisLeu 11
|||||
Db 359 GGAGACATTATTCTCCATCTT 379
|||||
RESULT 15
BH019424 425 bp DNA linear GSS 25-MAY-2001
LOCUS BH019424 L3443b.d.HygT7a.2 Leishmania major Friedlin Cosmid Genomic Library
DEFINITION Leishmania major genomic clone L3443b, genomic survey sequence.
ACCESSION BH019424
VERSION BH019424.1 GI:14198518
KEYWORDS GSS.
SOURCE Leishmania major
ORGANISM Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 425)
AUTHORS Myler,P.J., Vogt,C., Cawthra,J., Klacking,M., Marty,A., Mack,J.,
Munden,H., Nguyen,D., Robertson,L., Sisk,E., Fazelinia,G.,
Aggarwal,G., Nelson,S., Seyler,A., Worthey,E. and Stuart,K.
Leishmania major Friedlin Cosmid End Sequences
Unpublished (2000)
TITLE Other_GSS: L3443b.d.HygT7a.1
JOURNAL Contact: Myler PJ
COMMENT 4 Nickerson Street, Seattle, WA 98109-1651, USA
Tel: 206 284-8846
Fax: 206 284-0313
Email: mylerpj@sbrl.org
Seq primer: HygT7a

```

```

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 424)
AUTHORS Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
TITLE Functional annotation of a full-length Arabidopsis cDNA collection
JOURNAL Science 296 (5565), 141-145 (2002)
MEDLINE 21932900
PUBMED 11910074
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msek@tc.riken.go.jp
reversed clone; Please visit our web site
(http://pfgweb.gsc.riken.go.jp/) for further details.
FEATURES             Location/Qualifiers
     source           1..424
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL16-01-C17"
                     /lab_host="DH10B"
                     /clone_lib="RAFL16"
                     /note="Site 1: BamHI; Site 2: SalI; dark-grown"
ORIGIN
Alignment Scores: 550 Length: 424
Pred. No.: 7.00 Matches: 7
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 63.64% Gaps: 0
DB: 5
US-09-851-138C-155 (1-11) x BP597828 (1-424)
Qy 5 GlyAspIleLeuHisLeu 11
|||||
Db 359 GGAGACATTATTCTCCATCTT 379
|||||
RESULT 15
BH019424 425 bp DNA linear GSS 25-MAY-2001
LOCUS BH019424 L3443b.d.HygT7a.2 Leishmania major Friedlin Cosmid Genomic Library
DEFINITION Leishmania major genomic clone L3443b, genomic survey sequence.
ACCESSION BH019424
VERSION BH019424.1 GI:14198518
KEYWORDS GSS.
SOURCE Leishmania major
ORGANISM Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 425)
AUTHORS Myler,P.J., Vogt,C., Cawthra,J., Klacking,M., Marty,A., Mack,J.,
Munden,H., Nguyen,D., Robertson,L., Sisk,E., Fazelinia,G.,
Aggarwal,G., Nelson,S., Seyler,A., Worthey,E. and Stuart,K.
Leishmania major Friedlin Cosmid End Sequences
Unpublished (2000)
TITLE Other_GSS: L3443b.d.HygT7a.1
JOURNAL Contact: Myler PJ
COMMENT 4 Nickerson Street, Seattle, WA 98109-1651, USA
Tel: 206 284-8846
Fax: 206 284-0313
Email: mylerpj@sbrl.org
Seq primer: HygT7a

```

FEATURES Class: cosmid ends.
source Location/Qualifiers
1..425
/organism="Leishmania major"
/mol_type="genomic DNA"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="L3443b"
/lab_host="E. coli ED8767"
/clone_lib="Leishmania major Friedlin Cosmid Genomic Library"
/note="Vector: cLHYG; Site 1: BamHI; Genomic DNA from Leishmania major Friedlin was partially digested with Sau3AI, size selected, and ligated with BamHI-digested cLHYG cosmid vector DNA. 9216 clones were picked and arrayed. Library construction is described in Ivans et al., Genomics Research, 8:135-145 (1998). The cLHYG vector (Acc. No. CVU59231) is described in Ryan et al., Gene, 131:145-150 (1993)"

ORIGIN

Alignment Scores:
Pred. No.: 551 Length: 425
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 63.64% Indels: 0
DB: 8 Gaps: 0

US-09-851-138C-155 (1-11) x BH019424 (1-425)

Qy 5 GlyAspIleIleLeuHisLeu 11
Db 252 GGGGACATCATTCATCTC 272

Search completed: March 3, 2005, 21:58:16
Job time : 697.169 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 14:30:42 ; Search time 92.7333 Seconds
(without alignments)
829.870 Million cell updates/sec

Title: US-09-851-138C-174
Perfect score: 13
Sequence: 1 VRSGNTSCWIPV 13

Scoring table: QLIQGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8771383

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlp
-MODEL=frame+p2n.model
-O=/cpn2.1/USPTO.spool.p/US09851138/runat.28022005.120306.21457/app.query.fasta_1.1123
-DB=N_Geneseq_16Dec04 -OFMT=fastap -SUFFIX=olg.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=olg -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09851138 @CCN 1 1 1418 @runat.28022005.120306.21457 -NCPU=6 -ICPU=3
-NO MNAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn1990s.*
4: Geneseqn2000s.*
5: Geneseqn2001as.*
6: Geneseqn2001bs.*
7: Geneseqn2002as.*
8: Geneseqn2002bs.*
9: Geneseqn2003as.*
10: Geneseqn2003bs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	100.0	447	AAT27962	Aat27962 Hepatitis
2	10	76.9	576	AAQ83870	Aaq83870 Hepatitis
3	10	76.9	576	AAQ83873	Aaq83873 Hepatitis
4	10	76.9	576	AAQ83871	Aaq83871 Hepatitis
5	10	76.9	576	2 AAT16585	Aat16585 Hepatitis

6	10	76.9	576	2 AAT16584	Aat16584 Hepatitis
7	10	76.9	576	2 AAT16587	Aat16587 Hepatitis
8	10	76.9	2551	2 AAQ29630	Aaq29630 Hepatitis
9	10	76.9	2551	2 AAQ43891	Aaq43891 NANB hepa
10	10	76.9	9589	2 AAQ38218	Aaq38218 NANBH vir
11	9	69.2	356	2 AAT27949	Aat27949 Hepatitis
12	8	61.5	596	2 AAQ58819	Aaq58819 NANBH vir
13	8	61.5	596	12 ADJ81665	Adj81665 Non-A-non
14	8	61.5	1134	2 AAQ58821	Aaq58821 NANBH vir
15	8	61.5	1134	12 ADJ81667	Adj81667 Non-A-non
16	7	53.8	447	2 AAT27946	Aat27946 Hepatitis
17	7	53.8	447	2 AAT27948	Aat27948 Hepatitis
18	7	53.8	447	2 AAT27956	Aat27956 Hepatitis
19	7	53.8	574	2 AAQ78081	Aaq78081 Hepatitis
20	7	53.8	576	2 AAQ83849	Aaq83849 Hepatitis
21	7	53.8	576	2 AAQ83884	Aaq83884 Hepatitis
22	7	53.8	576	2 AAQ83848	Aaq83848 Hepatitis
23	7	53.8	576	2 AAQ83872	Aaq83872 Hepatitis
24	7	53.8	576	2 AAT16586	Aat16586 Hepatitis
25	7	53.8	576	2 AAT16598	Aat16598 Hepatitis
26	7	53.8	576	2 AAT16563	Aat16563 Hepatitis
27	7	53.8	576	2 AAT16562	Aat16562 Hepatitis
28	7	53.8	579	2 AAQ78103	Aaq78103 Hepatitis
29	7	53.8	582	4 AAH29583	Aah29583 Drosophil
30	7	53.8	775	8 ADA68681	Ada68681 Rice Gene
31	7	53.8	775	8 ADA68682	Ada68682 Rice Gene
32	7	53.8	957	2 AAQ78087	Aaq78087 Hepatitis
33	7	53.8	972	5 AAS14824	Aas14824 Human CDN
34	7	53.8	998	3 AAA37139	Aaa37139 Human UCP
35	7	53.8	998	4 AAF54487	Aaf54487 Primer #1
36	7	53.8	998	4 AAF92120	Aaf92120 Human PRO
37	7	53.8	998	6 ABS74440	Ab874440 Human CDN
38	7	53.8	998	8 ACA91226	ACA91226 Novel hum
39	7	53.8	998	8 ACD81603	ACD81603 Human CDN
40	7	53.8	998	8 ACA60425	ACA60425 Novel hum
41	7	53.8	998	8 ACA58872	ACA58872 CDNA enco
42	7	53.8	998	8 ACA64048	ACA64048 CDNA enco
43	7	53.8	998	8 ACA91312	ACA91312 CDNA enco
44	7	53.8	998	8 ACD45211	ACD45211 Human 88c
45	7	53.8	998	8 ACA93759	ACA93759 Human CDN

ALIGNMENTS

RESULT 1
AAT27962
ID AAT27962 standard; DNA; 447 BP.
XX
AC AAT27962;
XX
DT 11-MAR-1997 (first entry)
XX
DE Hepatitis C virus type 10a isolate NN98 bases 478-925.
XX
KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
KW PCR; primer; probe; antibody; infection; ss.
XX
OS Hepatitis C virus.
XX
PN WO9613590-A2.
XX
PD 09-MAY-1996.
XX
PF 23-OCT-1995; 95WO-EP004155.
PR 21-OCT-1994; 94EP-00870166.
PR 28-JUN-1995; 95EP-00870076.
XX
(INNO-) INNOGENETICS NV.
XX
PI Maertens G, Stuyver L;
XX
DR WPI; 1996-251460/25.

DR P-PSDB; AAR96551.
 XX
 PT Hepatitis C virus poly:nucleic acid unique to unidentified sub.type -
 PT used to develop probes and primers for new sub.types and vaccines to
 PT prevent and treat infection.
 XX
 PS Claim 6; Fig 3; 150pp; English.
 XX
 CC The sequences AAT27937-T27989 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f, 4a-
 CC 1, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l, 3g,
 CC 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
 CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the genome.
 CC This sequence represents nucleotides 478-925 from the HCV type 10a
 CC isolate NE98. The new HCV types were isolated from patients with chronic
 CC HCV from the Benelux countries, France, Cameroon and Vietnam, because of
 CC their aberrant reactivities. The RNA was extracted, cDNA synthesised and
 CC PCR amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions
 CC were sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences. The
 CC sequences were used to generate the peptides AAR96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect anti
 CC -HCV antibodies, for HCV typing or to prevent HCV infections
 XX
 SQ Sequence 447 BP; 82 A; 130 C; 114 G; 118 T; 0 U; 3 Other;
 Alignment Scores:
 Pred. No.: 3.75e-05 Length: 447
 Score: 13.00 Matches: 13
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0
 US-09-851-138C-174 (1-13) x AAT27962 (1-447)
 QY 1 ValArgSerGlyAanThrSerArgCysTrpIleProVal 13
 DB 211 GTACGCTGTGGCAATACATCAAGATGCTGGATCCCTGTG 249
 RESULT 2
 AAO83870
 ID AAO83870 standard; cDNA; 576 BP.
 XX
 AC AAO83870;
 XX
 DT 25-MAR-2003 (revised)
 DT 18-SEP-1995 (first entry)
 XX
 DE Hepatitis C virus envelope 1 gene cDNA isolate T2.
 XX
 KW Hepatitis C virus; HCV; non-A non-B; envelope 1 gene; isolate T2;
 KW diagnosis; vaccines; antibodies; antisera; gene inhibition; ss.
 XX
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 1..576
 FT /*tag= a
 XX
 PN WO9501442-A2.
 XX
 PD 12-JAN-1995.
 XX
 PF 28-JUN-1994; 94WO-US007320.
 XX
 PR 29-JUN-1993; 93US-00086428.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Bukh J, Miller RH, Purcell RH;
 XX
 DR WPI; 1995-061006/08.
 DR P-PSDB; AAR69662.
 XX
 PF Envelope 1 cDNAs of 51 hepatitis C virus isolates - and derived oligo-
 PF nucleotide(s), peptide(s) and proteins, used in diagnosis and in
 XX
 PT vaccines.
 XX
 PS Claim 1, Page 67; 186pp; English.
 XX
 CC AAO83873 encodes AAR69662 hepatitis C virus (HCV) envelope 1 (E1) protein

DR WPI; 1995-061006/08.
 DR P-PSDB; AAR69659.
 XX
 PT Envelope 1 cDNAs of 51 hepatitis C virus isolates - and derived oligo-
 PT nucleotide(s), peptide(s) and proteins, used in diagnosis and in
 XX
 PS vaccines.
 XX
 PS Claim 1; Page 65; 186pp; English.
 XX
 CC AAO83870 encodes AAR69659 hepatitis C virus (HCV) envelope 1 (E1) protein
 CC isolate T2, both can be used for the diagnosis of HCV infection, and in
 CC the prodn. of anti-HCV vaccines, antibodies and antisera. The cDNA may
 CC also be used to inhibit the expression of the HCV E1 gene. (Updated on 25
 CC -MAR-2003 to correct RN field.)
 XX
 SQ Sequence 576 BP; 104 A; 173 C; 175 G; 124 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.0767 Length: 576
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 76.92% Indels: 0
 DB: 2 Gaps: 0
 US-09-851-138C-174 (1-13) x AAO83870 (1-576)
 QY 4 GlyAanThrSerArgCysTrpIleProVal 13
 DB 124 GGAATACATCCGATGCTGGATACCGGTC 153
 RESULT 3
 AAO83873
 ID AAO83873 standard; cDNA; 576 BP.
 XX
 AC AAO83873;
 XX
 DT 25-MAR-2003 (revised)
 DT 18-SEP-1995 (first entry)
 XX
 DE Hepatitis C virus envelope 1 gene cDNA isolate US10.
 XX
 KW Hepatitis C virus; HCV; non-A non-B; envelope 1 gene; isolate US10;
 KW diagnosis; vaccines; antibodies; antisera; gene inhibition; ss.
 XX
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 1..576
 FT /*tag= a
 XX
 PN WO9501442-A2.
 XX
 PD 12-JAN-1995.
 XX
 PF 28-JUN-1994; 94WO-US007320.
 XX
 PR 29-JUN-1993; 93US-00086428.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Bukh J, Miller RH, Purcell RH;
 XX
 DR WPI; 1995-061006/08.
 DR P-PSDB; AAR69662.
 XX
 PF Envelope 1 cDNAs of 51 hepatitis C virus isolates - and derived oligo-
 PF nucleotide(s), peptide(s) and proteins, used in diagnosis and in
 XX
 PT vaccines.
 XX
 PS Claim 1, Page 67; 186pp; English.
 XX
 CC AAO83873 encodes AAR69662 hepatitis C virus (HCV) envelope 1 (E1) protein

CC isolate US10, both can be used for the diagnosis of HCV infection, and in
 CC the prodn. of anti-HCV vaccines, antibodies and antisera. The cDNA may
 CC also be used to inhibit the expression of the HCV E1 gene. (Updated on 25
 CC -MAR-2003 to correct PN field.)
 XX
 SQ Sequence 576 BP; 107 A; 169 C; 168 G; 132 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.0767 Length: 576
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 76.92% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x AAQ83873 (1-576)
 QY 4 GlyAenThrSerArgCysTrpIleProVal 13
 Db 124 GGAATACATCTCGGTCTGGATACCGGTC 153

RESULT 4
 AAQ83871
 ID AAQ83871 standard; cDNA; 576 BP.
 XX
 AC AAQ83871;
 XX
 XX 25-MAR-2003 (revised)
 DT 18-SEP-1995 (first entry)
 XX
 DE Hepatitis C virus envelope 1 gene cDNA isolate T4.

XX
 KW Hepatitis C virus; HCV; non-A non-B; envelope 1 gene; isolate T4;
 KW diagnosis; vaccines; antibodies; antisera; gene inhibition; ss.
 XX
 OS Hepatitis C virus.

XX
 FH Key Location/Qualifiers
 FT mat_peptide 1..576
 FT /*tag= a

XX WO9501442-A2.
 XX
 PD 12-JAN-1995.
 XX
 PF 28-JUN-1994; 94WO-US007320.
 XX
 PR 29-JUN-1993; 93US-00086428.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI Bukh J, Miller RH, Purcell RH;
 XX
 DR WPI: 1995-061006/08.
 DR P-PSDB; AAR69660.

XX
 PT Envelope 1 cDNAs of 51 hepatitis C virus isolates - and derived oligo-
 PT nucleotide(s), peptide(s) and proteins, used in diagnosis and in
 PT vaccines.

XX
 PS Claim 1; Page 65-66; 186pp; English.

XX
 CC AAQ83871 encodes AAR69660 hepatitis C virus (HCV) envelope 1 (E1) protein
 CC isolate T4, both can be used for the diagnosis of HCV infection, and in
 CC the prodn. of anti-HCV vaccines, antibodies and antisera. The cDNA may
 CC also be used to inhibit the expression of the HCV E1 gene. (Updated on 25
 CC -MAR-2003 to correct PN field.)
 XX

SQ Sequence 576 BP; 108 A; 171 C; 170 G; 127 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.0767 Length: 576
 Score: 10.00 Matches: 10

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 76.92% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x AAQ83871 (1-576)
 QY 4 GlyAenThrSerArgCysTrpIleProVal 13
 Db 124 GGAATACATCTCGGTCTGGATACCGGTT 153

RESULT 5
 AAT16585
 ID AAT16585 standard; cDNA; 576 BP.
 XX
 AC AAT16585;
 XX
 XX 30-SEP-1996 (first entry)
 DT
 DE Hepatitis C virus isolate T4 envelope 1 gene.

XX
 KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis; ss.
 XX
 OS Hepatitis C virus.

XX
 FH Key Location/Qualifiers
 FT CDS 1..576
 FT /*tag= a
 FT /product= "envelope-1 protein"
 FT /note= "does not contain start or stop codon"

XX WO9605315-A2.
 XX
 PD 22-FEB-1996.

PF 15-AUG-1995; 95WO-US010398.

PR 15-AUG-1994; 94US-00290665.

XX (USSH) US SEC DEPT HEALTH.

XX PI Bukh J, Miller RH, Purcell RH;

XX WPI: 1996-139709/14.

DR P-PSDB; AAR89531.

XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
 PT determine HCV genotype and as vaccines against HCV infection.

XX Claim 1; Page 95; 340pp; English.

XX
 CC AAT16559-T16609 are cDNAs encoding the E1 (envelope-1) protein of 51 HCV
 CC isolates. The isolated sequences are useful for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers are
 CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
 CC in vaccines for immunising against HCV infection. The proteins may also
 CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
 CC other mononuclear cells. The antibodies may be used in the prevention of
 CC HCV infection
 XX

SQ Sequence 576 BP; 108 A; 171 C; 170 G; 127 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.0767 Length: 576
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 76.92% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x AAT16585 (1-576)

QY 4 GlyAenThrSerArgCysTrpIleProVal 13
 Db 124 GGAATACATCTCGGTCTGGATACCGGT 153

RESULT 6
 AAT16584
 ID AAT16584 standard; cDNA; 576 BP.
 AC AAT16584;
 XX
 XX 30-SEP-1996 (first entry)
 XX Hepatitis C virus isolate T2 envelope 1 gene.
 DE HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 XX hepatitis; ss.
 KW
 XX Hepatitis C virus.
 OS
 XX Key Location/Qualifiers
 FH 1. .576
 FT CDS /tag= a
 FT /product= "envelope-1_protein"
 FT /note= "does not contain start or stop codon"
 XX
 PN WO9605315-A2.
 XX
 XX 22-FEB-1996.
 PD
 XX
 XX 15-AUG-1995; 95WO-US010398.
 PF
 XX 15-AUG-1994; 94US-00290665.
 PR
 XX (USSH) US SEC DEPT HEALTH.
 PA
 XX Bukh J, Miller RH, Purcell RH;
 PI WPI; 1996-139709/14.
 XX P-PSDB; AAR89530.
 DR
 XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
 PT determine HCV genotype and as vaccines against HCV infection.
 PT
 PS Claim 1; Page 94-95; 340pp; English.
 CC
 CC AAT16559-T16609 are cDNAs encoding the E1 (envelope-1) protein of 51 HCV
 CC isolates. The isolated sequences are useful for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers are
 CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
 CC in vaccines for immunising against HCV infection. The proteins may also
 CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
 CC other mononuclear cells. The antibodies may be used in the prevention of
 CC HCV infection
 XX
 SQ Sequence 576 BP; 104 A; 173 C; 175 G; 124 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.0767 Length: 576
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 76.92% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x AAT16584 (1-576)

QY 4 GlyAenThrSerArgCysTrpIleProVal 13
 Db 124 GGAATACATCCCGATCTGGATACCGGT 153

RESULT 7
 AAT16587
 ID AAT16587 standard; cDNA; 576 BP.

XX AAT16587;
 AC
 XX 30-SEP-1996 (first entry)
 DT
 XX Hepatitis C virus isolate US10 envelope 1 gene.
 DE
 XX HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis; ss.
 KW
 XX Hepatitis C virus.
 OS
 XX Key Location/Qualifiers
 FH 1. .576
 FT CDS /tag= a
 FT /product= "envelope-1_protein"
 FT /note= "does not contain start or stop codon"
 XX
 PN WO9605315-A2.
 XX
 XX 22-FEB-1996.
 PD
 XX
 XX 15-AUG-1995; 95WO-US010398.
 PF
 XX 15-AUG-1994; 94US-00290665.
 PR
 XX (USSH) US SEC DEPT HEALTH.
 PA
 XX Bukh J, Miller RH, Purcell RH;
 PI WPI; 1996-139709/14.
 XX P-PSDB; AAR89533.
 DR
 XX DNA and amino acid sequence of HCV envelope 1 and core proteins - used to
 PT determine HCV genotype and as vaccines against HCV infection.
 PT
 PS Claim 1; Page 96; 340pp; English.
 CC
 CC AAT16559-T16609 are cDNAs encoding the E1 (envelope-1) protein of 51 HCV
 CC isolates. The isolated sequences are useful for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers are
 CC also useful for HCV genotyping. Proteins encoded by the cDNAs can be used
 CC in vaccines for immunising against HCV infection. The proteins may also
 CC be used to detect antibodies against HCV in serum, saliva, lymphocytes or
 CC other mononuclear cells. The antibodies may be used in the prevention of
 CC HCV infection
 XX
 SQ Sequence 576 BP; 107 A; 169 C; 168 G; 132 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.0767 Length: 576
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 76.92% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x AAT16587 (1-576)

QY 4 GlyAenThrSerArgCysTrpIleProVal 13
 Db 124 GGAATACATCTCGGTCTGGATACCGGT 153

RESULT 8
 AAT29630
 ID AAT29630 standard; DNA; 2551 BP.
 AC AAT29630;
 XX
 XX 25-MAR-2003 (revised)
 DT 16-MAR-1993 (first entry)
 XX
 DE Hepatitis C virus HC-J6 5' region.

XX Non-A non-B hepatitis; NANBH; HCV; detection; diagnosis; screening; PCR;
 KW primer; polymerase chain reaction; ss.
 XX Hepatitis C virus.
 XX
 PN EP510952-A1.
 XX
 XX 28-OCT-1992.
 XX
 PF 23-APR-1992; 92EP-00303625.
 XX
 PR 26-APR-1991; 91JP-00191376.
 XX
 PA (IMMO) IMMUNO JAPAN INC.
 XX
 PI Okamoto H, Nakamura T;
 XX WPI; 1992-359137/44.
 DR
 PT Detection of non-A, non-B hepatitis virus - using new oligo-nucleotide
 PT primers with nucleotide sequences corresp. to part. of the viral RNA.
 XX
 PS Disclosure; Page 22; 54pp; English.
 XX
 CC This sequence represents the 5' region of hepatitis C virus RNA. The
 CC original sample was obtained from human and chimpanzee plasma. RNA was
 CC isolated from several samples and homology compared, and the respective
 CC sequence of about 1900 - 2500 nucleotides of the 5' terminus and 1100
 CC nucleotides of the 3' terminus determined. The 5' region (given) contains
 CC a non-coding region of at least 340 nucleotides and a region coding for
 CC the structural protein followed by a region coding for the non-structural
 CC protein (none actually detailed on the sequence given in the
 CC specification). When compared with the sequence of HCV disclosed in EP-
 CC 388232 this sequence showed homology of 72.5%. (Updated on 25-MAR-2003 to
 CC correct FN field.)
 XX
 SQ Sequence 2551 BP; 518 A; 779 C; 704 G; 550 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.283 Length: 2551
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 76.92% Indels: 0
 DB: 2 Gaps: 0
 US-09-851-138C-174 (1-13) x AAQ29630 (1-2551)
 QY 4 GlyAsnThrSerArgCysTrpIleProVal 13
 DB 1037 GGGATACATCTCGGTGCTGGATACCGGTC 1066
 RESULT 9
 AAQ43891
 ID AAQ43891 standard; cDNA to mRNA; 2551 BP.
 AC AAQ43891;
 XX
 XX 21-OCT-1993 (first entry)
 DE
 DE NANB hepatitis virus polynucleotide N-2551.
 KW Non-A, non-B; virus; polymerase chain reaction; detection; sensitive;
 KW specific; HCV; NANBH; ss.
 XX
 OS Non-A.
 OS non-B hepatitis virus.
 XX
 XX Key Location/Qualifiers
 FT 5'UTR 1..340
 FT /*tag= b
 FT /note= "from 5' terminal of NANBH virus RNA"
 FT

FT CDS 341..2551
 FT /*tag= a
 XX
 PN JP05091884-A.
 XX
 PD 16-APR-1993.
 XX
 PF 10-APR-1991; 91JP-00196175.
 XX
 PR 12-JUN-1990; 90JP-00153401.
 PR 08-NOV-1990; 90JP-00304405.
 XX
 PA (NAKA/) NAKAMURA T.
 XX
 DR WPI; 1993-199637/25.
 DR P-PSDB; AAR38281.
 XX
 PT Antigen related to non-A and non-B hepatitis virus - comprises non-
 PT translation region comprising 340 - 341 mols. of nucleotides, non-
 PT translation region comprising 1885 - 2551 mols. of nucleotides including
 PT region 1,149 and, etc.
 XX
 PS Claim 5; Page 23-24; 73pp; Japanese.
 XX
 CC The sequence is that of NANB hepatitis virus polynucleotide N-2551 which
 CC codes for a non-A, non-B (NANB) hepatitis virus gene HC-OM. The
 CC polypeptide it encodes may be used in a system for detecting NANB
 CC hepatitis. This method is highly specific and sensitive, and can detect
 CC NANB hepatitis virus which could not be detected by conventional methods
 XX
 SQ Sequence 2551 BP; 519 A; 778 C; 705 G; 549 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.283 Length: 2551
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 76.92% Indels: 0
 DB: 2 Gaps: 0
 US-09-851-138C-174 (1-13) x AAQ43891 (1-2551)
 QY 4 GlyAsnThrSerArgCysTrpIleProVal 13
 DB 1037 GGGATACATCTCGGTGCTGGATACCGGTC 1066
 RESULT 10
 AAQ38218
 ID AAQ38218 standard; cDNA to mRNA; 9589 BP.
 AC AAQ38218;
 XX
 XX 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 01-JUL-1993 (first entry)
 XX
 DE NANBH virus strain HC-J6 genome cDNA sequence.
 KW Non A non B hepatitis virus; amplification; HC-J1; HC-J8; plasma; ss.
 XX
 OS Non-A.
 OS non-B hepatitis virus.
 XX
 XX Key Location/Qualifiers
 FT CDS 341..9442
 FT /*tag= a
 XX
 PN EP532167-A2.
 XX
 PD 17-MAR-1993.
 XX
 PF 30-JUL-1992; 92EP-00306952.
 XX

PR 09-AUG-1991; 91JP-00287402.
 XX 05-DEC-1991; 91JP-00360441.
 XX (IMMO) IMMUNO JAPAN INC.
 XX Okamoto H, Nakamura T;
 XX WPI; 1993-087166/11.
 XX N-PSDB; AAR33538.
 XX Polynucleotide(s), polypeptide(s) and antibodies of NANBH virus - useful
 PT for detecting NANBH, as a vaccine and for screening blood samples.
 XX Claim 2; Page 27-32; 93pp; English.
 XX RNA was isolated from the plasma of human patients positive for NANBH
 CC virus (strain HC-J6) and was subjected to reverse transcription to
 CC produce cDNA. The resulting cDNA was amplified by PCR, and nucleic acid
 CC sequences determined by analysis of both clones from the cDNA library and
 CC clones obtd. by PCR amplification (36 clones in total). The NANBH HC-J6
 CC genome was found to contain an open reading frame encoding a polypeptide
 CC precursor of 3033 amino acid residues. See also AAR38172-221. (Updated on
 CC 25-MAR-2003 to correct PN field.) (Updated on 27-AUG-2003 to correct OS
 CC field.)
 XX SQ Sequence 9589 BP; 1968 A; 2820 C; 2635 G; 2166 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.903 Length: 9589
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 76.92% Indels: 0
 DB: 2 Gaps: 0
 US-09-851-138C-174 (1-13) x AAR38218 (1-9589)
 QY 4 GlyAsnThrSerArgCysTrpIleProVal 13
 DB 1037 GGGANATACATCTGGTGTGGATACCGGTC 1066
 RESULT 11
 AAT27949
 ID AAT27949 standard; DNA; 356 BP.
 XX AC AAT27949;
 XX DT 10-MAR-1997 (first entry)
 XX Hepatitis C virus type 2i isolate BNL6 bases 478-833.
 XX Hepatitis C virus; subtype; polymerase chain reaction; amplification;
 XX PCR; primer; probe; antibody; infection; ss.
 XX Hepatitis C virus.
 XX WO9613590-A2.
 XX 09-MAY-1996.
 XX 23-OCT-1995; 95WO-EF004155.
 XX 21-OCT-1994; 94EP-00870166.
 XX 28-JUN-1995; 95EP-00870076.
 XX (INNO-) INNOGENETICS NV.
 XX Maertens G, Stuyver L;
 XX WPI; 1996-251460/25.
 XX P-PSDB; AAR96538.
 XX Hepatitis C virus poly:nucleic acid unique to unidentified sub:type -

PT used to develop probes and primers for new sub:types and vaccines to
 XX prevent and treat infection.
 XX Claim 6; Fig 3; 150pp; English.
 XX The sequences AAT7937-T27989 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f, 4a-
 CC 1, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-1, 2k, 2l, 3g,
 CC 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5',
 CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the genome.
 CC This sequence represents nucleotides 478-833 from the HCV type 2i isolate
 CC BNL6. The new HCV types were isolated from patients with chronic HCV from
 CC the Benelux countries, France, Cameroon and Vietnam, because of their
 CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
 CC amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
 CC sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences. The
 CC sequences were used to generate the peptides AAR96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect anti
 CC -HCV antibodies, for HCV typing or to prevent HCV infections
 XX SQ Sequence 356 BP; 64 A; 107 C; 93 G; 92 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.593 Length: 356
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 69.23% Indels: 0
 DB: 2 Gaps: 0
 US-09-851-138C-174 (1-13) x AAT27949 (1-356)
 QY 5 AsnThrSerArgCysTrpIleProVal 13
 DB 223 AACACCTCCGCTGGTGGATACCGGTC 249
 RESULT 12
 AAR958819
 ID AAR958819 standard; cDNA; 596 BP.
 XX AC AAR958819;
 XX DT 27-AUG-2003 (revised)
 DT 24-NOV-1994 (first entry)
 XX NANBH virus gene fragment #6.
 XX Antigen; structural; non-structural; non A non B hepatitis virus; NANBH;
 KW NANBH; patient; plasma; diagnosis; detection; carrier; ss.
 XX Non-A.
 OS non-B hepatitis virus.
 XX Key Location/Qualifiers
 FT 1. .594
 FT /*tag= a
 FT /product= "NANBH protein fragment"
 XX JP06070778-A.
 XX 15-MAR-1994.
 XX 01-JUN-1993; 93JP-00156087.
 XX 10-JUL-1992; 92JP-00207391.
 XX (TOKR-) ZH TOKYO RINSHO IGAKU SOGO KENKYUSHO.
 PA (SANW) SANWA KAGAKU KENKYUSHO CO.
 PA (TOFU) TONEN CORP.
 PA (KOKU-) KOKUSAI SHIYAKU KK.
 XX

DR WPI; 1994-128677/16.
 DR P-PSDB; AAR50074.
 XX Nucleic acid fragment coding non-A non-B hepatitis virus antigen - useful
 PT in diagnosis of NANB patient and detection of virus carrier.
 XX Claim 12; Page 21; 37pp; Japanese.
 XX The sequences given in AAQ58814-27 encode antigens of structural and non-
 CC structural regions of non A non B hepatitis virus (NANBH). These
 CC sequences were derived from the plasma of a NANBH patient by recombinant
 CC DNA techniques. These fragments are useful for the diagnosis of NANBH
 CC patients and the detection of NANBH carriers. (Updated on 27-AUG-2003 to
 CC correct OS field.)
 XX SQ Sequence 596 BP; 101 A; 183 C; 175 G; 137 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 11 Length: 596
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 61.54% Indels: 0
 DB: 2 Gaps: 0
 US-09-851-138C-174 (1-13) x AAQ58819 (1-596)
 QY 6 ThrSerArgCysTrpIleProVal 13
 DB 328 ACGTCACGGTCTGGATACCGGTC 351
 RESULT 13
 ADJ81665
 ID ADJ81665 standard; cDNA; 596 BP.
 XX AC ADJ81665;
 XX DT 06-MAY-2004 (first entry)
 DE Non-A-non-B hepatitis antigen cDNA sequence SeqID6.
 KW non-A-non-B type hepatitis virus antigen; recombinant technique;
 KW hepatitis C virus infection; Gene; ss.
 XX OS Hepatitis C virus.
 FH Key Location/Qualifiers
 FT CDS 1..594
 FT /tags a
 FT /product= "Non-A-non-B type hepatitis antigen"
 FT /partial
 FT /note= "No start or stop codon"
 XX JP2004000151-A.
 XX PD 08-JAN-2004.
 XX PF 24-FEB-2003; 2003JP-00046384.
 XX PR 10-JUL-1992; 92JP-00207391.
 XX PR 01-JUN-1993; 93JP-00156087.
 XX PA (KOKU-) KOKUSAI SHIYAKU KK.
 XX WPI; 2004-085214/09.
 DR P-PSDB; ADJ81679.
 XX Novel nucleic acid fragment which codes for non-A-non-B type hepatitis
 PT virus antigen, useful for diagnosing hepatitis C virus infection in a
 PT patient.
 XX PS Disclosure; SEQ ID NO 6; 59pp; Japanese.

CC This invention relates to a novel nucleic acid fragment containing a
 CC nucleotide sequence which codes for non-A-non-B type hepatitis virus
 CC antigen which has a fully defined sequence of 273 or 330 amino acids as
 CC given in the specification. The invention is useful for producing non-A-
 CC non-B type hepatitis virus antigen by recombinant techniques. The
 CC invention may therefore be useful for diagnosing hepatitis C virus
 CC infection in a patient and thus helping in prevention of the disease. The
 CC invention allows effective detection of non-A-non-B hepatitis patients.
 CC The present sequence is that of a cDNA sequence of the invention which
 CC encodes a non-A-non-B type hepatitis virus antigen.
 XX SQ Sequence 596 BP; 101 A; 183 C; 175 G; 137 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 11 Length: 596
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 61.54% Indels: 0
 DB: 12 Gaps: 0
 US-09-851-138C-174 (1-13) x ADJ81665 (1-596)
 QY 6 ThrSerArgCysTrpIleProVal 13
 DB 328 ACGTCACGGTCTGGATACCGGTC 351
 RESULT 14
 AAQ58821
 ID AAQ58821 standard; cDNA; 1134 BP.
 XX AC AAQ58821;
 XX DT 27-AUG-2003 (revised)
 DT 24-NOV-1994 (first entry)
 XX DE NANBH virus gene fragment #8.
 KW Antigen; structural; non-structural; non A non B hepatitis virus; NANBH;
 KW NANBH; patient; plasma; diagnosis; detection; carrier; ss.
 XX OS Non-A.
 OS non-B hepatitis virus.
 XX JP06070778-A.
 XX PN 15-MAR-1994.
 PD 01-JUN-1993; 93JP-00156087.
 XX PR 10-JUL-1992; 92JP-00207391.
 XX PA (TOXR-) ZH TOKYO RINSHO IGAKU SOGO KENKYUSHO.
 PA (SANW) SANWA KAGAKU KENKYUSHO CO.
 PA (TOFU) TONEN CORP.
 PA (KOKU-) KOKUSAI SHIYAKU KK.
 XX WPI; 1994-128677/16.
 DR P-PSDB; AAR58821.
 XX Nucleic acid fragment coding non-A non-B hepatitis virus antigen - useful
 PT in diagnosis of NANB patient and detection of virus carrier.
 XX Claim 16; Page 23-24; 37pp; Japanese.
 XX The sequences given in AAQ58814-27 encode antigens of structural and non-
 CC structural regions of non A non B hepatitis virus (NANBH). These
 CC sequences were derived from the plasma of a NANBH patient by recombinant
 CC DNA techniques. These fragments are useful for the diagnosis of NANBH
 CC patients and the detection of NANBH carriers. (Updated on 27-AUG-2003 to
 CC correct OS field.)
 XX SQ Sequence 1134 BP; 214 A; 347 C; 317 G; 256 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 19,3 Length: 1134
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 61.54% Indels: 0
 DB: 2 Gaps: 0
 US-09-851-138C-174 (1-13) x AAQ58821 (1-1134)

QY 6 ThrSerArgCysTrpIleProVal 13
 |||||
 Db 328 ACGTCACGGTGTGGATACCGGTC 351

RESULT 15

ADJ81667
 ID ADJ81667 standard; cDNA; 1134 BP.

XX AC ADJ81667;

XX XX 06-MAY-2004 (first entry)

XX DE Non-A-non-B hepatitis antigen cDNA sequence SeqID8.

XX KW non-A-non-B type hepatitis virus antigen; recombinant technique;
 XX KW hepatitis C virus infection; gene; ss.

XX OS Hepatitis C virus.

XX FH Key Location/Qualifiers

FT CDS 1..1134
 FT /*tag= a
 FT /product= "Non-A-non-B type hepatitis antigen"
 FT /partial
 FT /note= "No start or stop codon"

XX PN JP2004000151-A.

XX XX 08-JAN-2004.

XX XX 24-FEB-2003; 2003JP-00046384.

XX PR 10-JUL-1992; 92JP-00207391.

XX PR 01-JUN-1993; 93JP-00156087.

XX PA (KOKU-) KOKUSAI SHIYAKU KK.

XX XX WPI; 2004-085214/09.

DR P-PSDB; ADJ81681.

XX PT Novel nucleic acid fragment which codes for non-A-non-B type hepatitis
 PT virus antigen, useful for diagnosing hepatitis C virus infection in a
 PT patient.

XX PS Disclosure; SEQ ID NO 8; 59pp; Japanese.

XX CC This invention relates to a novel nucleic acid fragment containing a
 CC nucleotide sequence which codes for non-A-non-B type hepatitis virus
 CC antigen which has a fully defined sequence of 273 or 330 amino acids as
 CC given in the specification. The invention is useful for producing non-A-
 CC non-B type hepatitis virus antigen by recombinant techniques. The
 CC invention may therefore be useful for diagnosing hepatitis C virus
 CC infection in a patient and thus helping in prevention of the disease. The
 CC invention allows effective detection of non-A-non-B hepatitis patients.
 CC The present sequence is that of a cDNA sequence of the invention which
 CC encodes a non-A-non-B type hepatitis virus antigen.

XX SQ Sequence 1134 BP; 214 A; 347 C; 317 G; 256 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 19,3 Length: 1134
 Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 61.54% Indels: 0
 DB: 12 Gaps: 0
 US-09-851-138C-174 (1-13) x ADJ81667 (1-1134)

QY 6 ThrSerArgCysTrpIleProVal 13
 |||||
 Db 328 ACGTCACGGTGTGGATACCGGTC 351

Search completed: March 3, 2005, 16:26:10
 Job time : 94.7333 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:54:32 ; Search time 26.8 Seconds

(without alignments)
793.716 Million cell updates/sec

Title: US-09-851-138c-174

Perfect score: 13

Sequence: 1 VRSGNTRCWPV 13

Scoring table:

QXgap 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2396881

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Oa/cgn2_1/USPTO.epool_p/US09851138/runat.28022005.120307.21495/app.query.fasta_1.1123
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=olg.rn1 -MINWATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09851138 @CIGN 1.1.249 @runat.28022005.120307.21495 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	100.0	447	3	US-08-836-075A-51
2	10	76.9	277	2	US-08-634-797-42
3	10	76.9	306	2	US-08-634-797-17
4	10	76.9	333	2	US-08-634-797-21
5	10	76.9	333	2	US-08-634-797-31
6	10	76.9	576	1	US-08-086-428B-26
7	10	76.9	576	1	US-08-086-428B-27
8	10	76.9	576	1	US-08-086-428B-29
9	10	76.9	576	2	US-08-468-570-26
10	10	76.9	576	2	US-08-468-570-27
11	10	76.9	576	2	US-08-468-570-29
12	10	76.9	576	2	US-08-290-665A-26

13	10	76.9	576	2	US-08-290-665A-27	Sequence 27, Appl
14	10	76.9	576	2	US-08-290-665A-29	Sequence 29, Appl
15	10	76.9	576	4	US-08-466-601A-26	Sequence 26, Appl
16	10	76.9	576	4	US-08-466-601A-27	Sequence 27, Appl
17	10	76.9	576	4	US-08-466-601A-29	Sequence 29, Appl
18	10	76.9	576	5	PCT-US95-10398-26	Sequence 26, Appl
19	10	76.9	576	5	PCT-US95-10398-27	Sequence 27, Appl
20	10	76.9	576	5	PCT-US95-10398-29	Sequence 29, Appl
21	10	76.9	9589	1	US-07-925-695-1	Sequence 1, Appl
22	10	76.9	9589	1	US-07-925-695-2	Sequence 2, Appl
23	9	69.2	356	3	US-08-836-075A-25	Sequence 25, Appl
24	7	53.8	333	2	US-08-634-797-25	Sequence 25, Appl
25	7	53.8	447	3	US-08-836-075A-19	Sequence 19, Appl
26	7	53.8	447	3	US-08-836-075A-23	Sequence 23, Appl
27	7	53.8	447	3	US-08-836-075A-39	Sequence 39, Appl
28	7	53.8	574	4	US-09-878-281A-120	Sequence 120, Appl
29	7	53.8	576	1	US-08-086-428B-4	Sequence 4, Appl
30	7	53.8	576	1	US-08-086-428B-5	Sequence 5, Appl
31	7	53.8	576	1	US-08-086-428B-28	Sequence 28, Appl
32	7	53.8	576	1	US-08-086-428B-40	Sequence 40, Appl
33	7	53.8	576	2	US-08-468-570-4	Sequence 4, Appl
34	7	53.8	576	2	US-08-468-570-5	Sequence 5, Appl
35	7	53.8	576	2	US-08-468-570-28	Sequence 28, Appl
36	7	53.8	576	2	US-08-468-570-40	Sequence 40, Appl
37	7	53.8	576	2	US-08-290-665A-4	Sequence 4, Appl
38	7	53.8	576	2	US-08-290-665A-5	Sequence 5, Appl
39	7	53.8	576	2	US-08-290-665A-28	Sequence 28, Appl
40	7	53.8	576	2	US-08-290-665A-40	Sequence 40, Appl
41	7	53.8	576	4	US-08-466-601A-4	Sequence 4, Appl
42	7	53.8	576	4	US-08-466-601A-5	Sequence 5, Appl
43	7	53.8	576	4	US-08-466-601A-28	Sequence 28, Appl
44	7	53.8	576	4	US-08-466-601A-40	Sequence 40, Appl
45	7	53.8	576	5	PCT-US95-10398-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-08-836-075A-51
; Sequence 51, Application US/09836075A
; Patent No. 6180768

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; APPLICANT: STUYVER, LIEVEN

; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES

; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC

; TITLE OF INVENTION: AGENTS

; NUMBER OF SEQUENCES: 207

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ARNOLD, WHITE & DURKEE

; STREET: P.O. BOX 4433

; CITY: HOUSTON

; STATE: TEXAS

; COUNTRY: USA

; ZIP: 77210-4433

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Microsoft Word 6.0 / ASCII text output

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/836,075A

; FILING DATE: 21 Apr 1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/EP95/04155

; FILING DATE: 23 Oct 1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 94870166.9

; FILING DATE: 21 Oct 1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 95870076.7

; FILING DATE: 28 Jun 1995

; ATTORNEY/AGENT INFORMATION:

```

; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-836-075A-51

Alignment Scores:
Pred. No.: 3,688-06 Length: 447
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-851-138C-174 (1-13) x US-08-836-075A-51 (1-447)
Qy 1 ValArgSerGlyAsnThrSerArgCysTrpIleProVal 13
Db 211 GTACGCTCTGGCAATACATCAAGATGCTGGATCCCTGTG 249

RESULT 2
US-08-634-797-42
; Sequence 42, Application US/08634797
; Patent No. 5851759
; GENERAL INFORMATION:
; APPLICANT: WEINER, AMY J.
; TITLE OF INVENTION: HETERODUPLEX TRACKING ASSAY (HTA) FOR
; TITLE OF INVENTION: GENOTYPING HCV
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street - R440
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/634,797
; FILING DATE: 19-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 1226.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-3274
; TELEFAX: (510) 655-3542
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-634-797-42

Alignment Scores:
Pred. No.: 0,00523 Length: 277
Score: 10.00 Matches: 10
```

```

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x US-08-634-797-42 (1-277)
Qy 4 GlyAsnThrSerArgCysTrpIleProVal 13
Db 140 GGGAAATACATCTCGGTGCTGGATACCGGTC 169

RESULT 3
US-08-634-797-17
; Sequence 17, Application US/08634797
; Patent No. 5851759
; GENERAL INFORMATION:
; APPLICANT: WEINER, AMY J.
; TITLE OF INVENTION: HETERODUPLEX TRACKING ASSAY (HTA) FOR
; TITLE OF INVENTION: GENOTYPING HCV
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street - R440
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/634,797
; FILING DATE: 19-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 1226.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-3274
; TELEFAX: (510) 655-3542
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-634-797-17

Alignment Scores:
Pred. No.: 0,00571 Length: 306
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x US-08-634-797-17 (1-306)
Qy 4 GlyAsnThrSerArgCysTrpIleProVal 13
Db 73 GGGAAATACATCTCGGTGCTGGATACCGGTC 102

RESULT 4
US-08-634-797-21
; Sequence 21, Application US/08634797
; Patent No. 5851759
; GENERAL INFORMATION:
; APPLICANT: WEINER, AMY J.
```

;/ TITLE OF INVENTION: HETERO DUPLICATION TRACKING ASSAY (HTA) FOR
;/ TITLE OF INVENTION: GENOTYPING HCV
;/ NUMBER OF SEQUENCES: 52
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Chiron Corporation
;/ STREET: 4560 Horton Street - R440
;/ CITY: Emeryville
;/ STATE: California
;/ COUNTRY: USA
;/ ZIP: 94608-2916
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: Patent In Release #1.0, Version #1.30
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/634,797
;/ FILING DATE: 19-APR-1996
;/ CLASSIFICATION: 435
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Harbin, Alisa A.
;/ REGISTRATION NUMBER: 33,895
;/ REFERENCE/DOCKET NUMBER: 1226.001
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (510) 601-3274
;/ TELEFAX: (510) 655-3542
;/ TELEX: N/A
;/ INFORMATION FOR SEQ ID NO: 21:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 333 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: DNA (genomic)
;/ US-08-634-797-21
;/
Alignment Scores:
Pred. No.: 0.00616 Length: 333
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 2 Gaps: 0
;/
US-09-851-138C-174 (1-13) x US-08-634-797-21 (1-333)
QY 4 GlyAenThSrArGcYstTlPleProVal 13
DB 115 GGAAATACATCTCGGTGCTGGATACCGGTC 144
;/
RESULT 5
US-08-634-797-31
;/ Sequence 31, Application US/08634797
;/ Patent No. 5851759
;/ GENERAL INFORMATION:
;/ APPLICANT: WEINER, AMY J.
;/ TITLE OF INVENTION: HETERO DUPLICATION TRACKING ASSAY (HTA) FOR
;/ TITLE OF INVENTION: GENOTYPING HCV
;/ NUMBER OF SEQUENCES: 52
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Chiron Corporation
;/ STREET: 4560 Horton Street - R440
;/ CITY: Emeryville
;/ STATE: California
;/ COUNTRY: USA
;/ ZIP: 94608-2916
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: Patent In Release #1.0, Version #1.30
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/634,797

;/ FILING DATE: 19-APR-1996
;/ CLASSIFICATION: 435
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Harbin, Alisa A.
;/ REGISTRATION NUMBER: 33,895
;/ REFERENCE/DOCKET NUMBER: 1226.001
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (510) 601-3274
;/ TELEFAX: (510) 655-3542
;/ TELEX: N/A
;/ INFORMATION FOR SEQ ID NO: 31:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 333 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: DNA (genomic)
;/ US-08-634-797-31
;/
Alignment Scores:
Pred. No.: 0.00616 Length: 333
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 2 Gaps: 0
;/
US-09-851-138C-174 (1-13) x US-08-634-797-31 (1-333)
QY 4 GlyAenThSrArGcYstTlPleProVal 13
DB 115 GGAAATACATCTCGGTGCTGGATACCGGTC 144
;/
RESULT 6
US-08-086-428B-26
;/ Sequence 26, Application US/08086428B
;/ Patent No. 5514539
;/ GENERAL INFORMATION:
;/ APPLICANT: BUKH, J., MILLER, R.H. AND
;/ APPLICANT: PURCELL, R.H.
;/ TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
;/ TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE
;/ TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE
;/ TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN
;/ TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES
;/ NUMBER OF SEQUENCES: 159
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: MORGAN & FINNEGAN
;/ STREET: 345 PARK AVENUE
;/ CITY: NEW YORK
;/ STATE: NEW YORK
;/ COUNTRY: USA
;/ ZIP: 10154
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: FLOPPY DISK
;/ COMPUTER: IBM PC COMPATIBLE
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: WORDPERFECT 5.1
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/086.428B
;/ FILING DATE: 29-JUN-1993
;/ CLASSIFICATION: 435
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: RICHARD W. BORK
;/ REGISTRATION NUMBER: 36,459
;/ REFERENCE/DOCKET NUMBER: 2026-4070
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (212) 758-4800
;/ TELEFAX: (212) 751-6849
;/ TELEX: 421792
;/ INFORMATION FOR SEQ ID NO: 26:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 576 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: T2
US-08-086-428B-26

Alignment Scores:
Pred. No.: 0.01 Length: 576
Score: 10.00 Matches: 10
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Indels: 0
DB: 1
Gaps: 0

US-09-851-138C-174 (1-13) x US-08-086-428B-26 (1-576)

QY 4 GlyAsnThrSerArgCysTrpIleProVal 13
DB 124 GGAATACATCCGATCGTGATACCGGTC 153

RESULT 7
; Sequence 27, Application US/08086428B
; Patent No. 5514539
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE
; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE
; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN
; NUMBER OF SEQUENCES: 159
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/086,428B
; FILING DATE: 29-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4070
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: T4
US-08-086-428B-27

Alignment Scores:
Pred. No.: 0.01 Length: 576
Score: 10.00 Matches: 10
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Indels: 0
DB: 1
Gaps: 0

```

```

Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Indels: 0
DB: 1
Gaps: 0

US-09-851-138C-174 (1-13) x US-08-086-428B-27 (1-576)

QY 4 GlyAsnThrSerArgCysTrpIleProVal 13
DB 124 GGAATACATCTCGGTGCTGGATACCGGTT 153

RESULT 8
US-08-086-428B-29
; Sequence 29, Application US/08086428B
; Patent No. 5514539
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE
; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE
; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN
; NUMBER OF SEQUENCES: 159
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/086,428B
; FILING DATE: 29-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4070
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: 10
US-08-086-428B-29

Alignment Scores:
Pred. No.: 0.01 Length: 576
Score: 10.00 Matches: 10
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Indels: 0
DB: 1
Gaps: 0

US-09-851-138C-174 (1-13) x US-08-086-428B-29 (1-576)

QY 4 GlyAsnThrSerArgCysTrpIleProVal 13
DB 124 GGAATACATCTCGGTGCTGGATACCGGTC 153

```

RESULT 9

US-08-468-570-26
; Sequence 26, Application US/08468570
; Patent No. 5871962
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE
; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE
; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN
; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 159
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,570
; FILING DATE: 6-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29-JUN-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4070US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792

INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: T2
US-08-468-570-26

Alignment Scores:
Pred. No.: 0.01 Length: 576
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x US-08-468-570-26 (1-576)

QY 4 GlyAsnThrSerArgCysTrpIleProVal 13
Db 124 GGAATACATCCGATCGTGGATACCGGTC 153

RESULT 10

US-08-468-570-27
; Sequence 27, Application US/08468570
; Patent No. 5871962
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.

; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE
; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE
; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN
; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 159
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,570
; FILING DATE: 6-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29-JUN-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4070US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792

INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: T4
US-08-468-570-27

Alignment Scores:
Pred. No.: 0.01 Length: 576
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x US-08-468-570-27 (1-576)

QY 4 GlyAsnThrSerArgCysTrpIleProVal 13
Db 124 GGAATACATCCGATCGTGGATACCGGTT 153

RESULT 11

US-08-468-570-29
; Sequence 29, Application US/08468570
; Patent No. 5871962
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE
; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE
; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN
; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 159
; CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,570
FILING DATE: 6-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29-JUN-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4070US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 576 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: 10
US-08-468-570-29

Alignment Scores:
Pred. No.: 0.01 Length: 576
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x US-08-468-570-29 (1-576)

QY 4 GlyAthrSerArgCysTlIleProVal 13
Db 124 GGAATACATCTCGGTGCTGGATACCGGTC 153

RESULT 12
US-08-290-665A-26
; Sequence 26, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 576 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: T2
US-08-290-665A-26

Alignment Scores:
Pred. No.: 0.01 Length: 576
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x US-08-290-665A-26 (1-576)

QY 4 GlyAthrSerArgCysTlIleProVal 13
Db 124 GGAATACATCTCGGTGCTGGATACCGGTC 153

RESULT 13
US-08-290-665A-27
; Sequence 27, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 576 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: T4
US-08-290-665A-27

Alignment Scores:
Pred. No.: 0.01 Length: 576
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x US-08-290-665A-27 (1-576)

Qy 4 GlyAanThrSerArgCysTrpIleProVal 13
Db 124 GGAATACATCTCGGTGCTGGATACCGGTT 153

RESULT 14

US-08-290-665A-29
Sequence 29, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 576 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: US10
US-08-290-665A-29
Alignment Scores:
Pred. No.: 0.01 Length: 576
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 76.92% Indels: 0
DB: 2 Gaps: 0
US-09-851-138C-174 (1-13) x US-08-290-665A-29 (1-576)
Qy 4 GlyAanThrSerArgCysTrpIleProVal 13
Db 124 GGAATACATCTCGGTGCTGGATACCGGTC 153
RESULT 15
US-08-466-601A-26
Sequence 26, Application US/08466601A
Patent No. 6572864
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE
TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE
TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN
TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,601A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29-JUN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4070US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 576 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: T2
US-08-466-601A-26
Alignment Scores:
Pred. No.: 0.01 Length: 576

Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 76.92% Indels: 0
 DB: 4 Gaps: 0

US-09-851-138C-174 (1-13) x US-08-466-601A-26 (1-576)

Qy 4 GlyAsnThrSerArgCysTrpIleProVal 13
 |||||
 Db 124 GGAAATACATCCCGATGCTGGATACCGGTC 153

Search completed: March 3, 2005, 22:05:16
 Job time : 27.8 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:43:48 ; Search time 819.2 Seconds
(without alignments)
604.047 Million cell updates/sec

Title: US-09-851-138C-174
Perfect score: 13
Sequence: 1 VRSGNTSRCWIPV 13

Scoring table: 4OUTGGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Word size: 1

Total number of hits satisfying chosen parameters: 68473426

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlp
-O=/cpn2_1/USPTO spool_p/US09851138/runat_28022005_120306_21476/app query.fasta_1.1123
-DB=EST-QFMT=fastap SUPFIX=olig.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=b1cs -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:

1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gsa1:*
9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	69.2	549	BQ039335	BQ039335 gd09a08.y
2	9	69.2	688	BJ168506	BJ168506 BJ168506
3	8	61.5	393	AQ125006	AQ125006 HS 2163.A
4	8	61.5	731	BF611925	BF611925 df13b03.Y
5	8	61.5	749	CE452932	CE452932 tigr-gss-
6	8	61.5	815	CC336597	CC336597 OQOAX72TV
7	8	61.5	894	CG167588	CG167588 PUFLA51TB
8	8	61.5	904	CG167590	CG167590 PUFLA51TD
9	8	61.5	968	CL490252	CL490252 SAIL_534_

10	7	53.8	164	4	BQ052637	BQ052637 BJ052637
11	7	53.8	179	2	BF866593	BF866593 RC6-TN007
12	7	53.8	201	6	CB460281	CB460281 720152 MA
13	7	53.8	227	8	B93974	B93974 CIT-HSP-216
14	7	53.8	271	6	CA525295	CA525295 KS12053A0
15	7	53.8	285	5	BW202238	BW202238 BW202238
16	7	53.8	298	6	CA226543	CA226543 SCRLFL300
17	7	53.8	305	8	BH387061	BH387061 AG-ND-134
18	7	53.8	319	1	AV750690	AV750690 AV750690
19	7	53.8	319	5	BQ343995	BQ343995 IL3-NT010
20	7	53.8	323	9	CG836579	CG836579 ZMBBRC021
21	7	53.8	335	6	CA226614	CA226614 SCRLFL300
22	7	53.8	355	8	AQ445235	AQ445235 GSSTC0152
23	7	53.8	360	2	BE711547	BE711547 QV2-HT089
24	7	53.8	371	8	AQ090884	AQ090884 GSSTC0100
25	7	53.8	372	1	AI057833	AI057833 TENU1924
26	7	53.8	376	6	CA621213	CA621213 w11n.pk00
27	7	53.8	379	7	CF336988	CF336988 JMT--07-E
28	7	53.8	384	8	AZ050069	AZ050069 GSSTC1104
29	7	53.8	388	1	AV655181	AV655181 AV655181
30	7	53.8	388	8	AQ06416	AQ06416 GSSTC0446
31	7	53.8	391	2	BE772466	BE772466 RC2-FT012
32	7	53.8	393	2	BE772458	BE772458 RC2-FT012
33	7	53.8	394	8	AQ090293	AQ090293 GSSTC0780
34	7	53.8	398	8	AQ090131	AQ090131 GSSTC0756
35	7	53.8	398	8	AQ444042	AQ444042 GSSTC0884
36	7	53.8	401	8	AQ090416	AQ090416 GSSTC0988
37	7	53.8	405	2	BF287059	BF287059 EST451650
38	7	53.8	406	7	CF571553	CF571553 MCS015B07
39	7	53.8	407	1	AI055786	AI055786 SMOVL3CAN
40	7	53.8	407	2	BE772467	BE772467 RC2-FT012
41	7	53.8	408	7	CF314819	CF314819 HD--03-H2
42	7	53.8	408	8	AQ609959	AQ609959 HS 5089.A
43	7	53.8	409	2	BE772459	BE772459 RC2-PT012
44	7	53.8	409	8	AQ090398	AQ090398 GSSTC0583
45	7	53.8	410	1	AI328819	AI328819 a6hlone.f

ALIGNMENTS

RESULT 1
BQ039335

LOCUS
DEFINITION

ACCESSION
VERSION

KEYWORDS
SOURCE

ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

BQ039335 549 bp mRNA linear EST 17-APR-2002
gd09a08.y1 Moss EST library PPS Physcomitrella patens cDNA clone
PEP SOURCE_ID:PPS30116 5', mRNA sequence.

BQ039335
BQ039335.1 GI:19778637

EST.
Physcomitrella patens

Physcomitrella patens
Eukaryota; Viridiplantae;

Bryopsida; Funariidae; Funariaceae; Physcomitrella.

1 (bases 1 to 549)

Quatrano, R., Bashardes, S., Cove, D., Cumming, A., Knight, C.,
Clifton, S., Barra, M., Hillier, L., Pape, D., Martin, J., Wylie, T.,
Underwood, K., Theising, B., Allen, M., Bowers, Y., Person, B.,
Swaller, T., Steptoe, M., Gibbons, M., Harvey, N., Ritter, E.,
Jackson, Y., McCann, R., Waterston, R. and Wilson, R.

Leeds/Wash U Moss EST Project
Unpublished (1999)

Other ESTs: gd09a08.x1
Contact: Ralph Quatrano

Leeds/Wash U Moss EST Project
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800

Fax: 314 286 1810
Email: est@wustl.edu

Libraries were constructed by Dr. Stavros Bashardes as part of the
Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and

Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a

clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)

Seq primer: -40RP from Gibco
High quality sequence stop: 422.

FEATURES

source

1. 549
/organism="Physcomitrella patens"
/mol_type="mRNA"
/db_xref="taxon:3218"
/clone="PEP SOURCE ID:PPS30116"
/dev_stage="protonemata, 7day old untreated"
/lab_host="E.coli DH10b"
/clone_lib="Moss EST library PPS"
/note="Vector: pBluescript SK-; Site_1: XhoI; Site_2: EcoRI; Library constructed by Stavros Bashirades and re-arranged by A. Cuming & Honglin Rong. Construction of the cDNA library was carried out using Statagene's 'UniZAP' - cDNA synthesis kit, to ligate cDNA directionally in UniZAP XR vector arms. The vector is designed containing the pBluescript sequence as well as the lambda DNA and cDNA is cloned in the EcoRI and XhoI sites in the pBluescript sequence. The vector was then packaged using Gold gigapackaging extracts, propagated in XL-IBLue MRF cells and amplified. The library was excised by mass excision using Stratagene's Mass excision kit to infect SOLR cells with phagemids and ampicillin resistant transformants selected. Approximately 1,000,000 colonies were grown and recovered by using Quiagen midi prep kit.2 micro grams of plasmid DNA were used to transform DH10b cells by electroporation. Clones corresponding to abundant transcripts were identified by colony hybridization and eliminated from the library, be rearraying. This library is non-directionally cloned."

Alignment Scores:
Pred. No.: 8.58 Length: 549
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 69.23% Indels: 0
DB: 5 Gaps: 0

ORIGIN

US-09-851-138C-174 (1-13) x BQ039335 (1-549)
Qy 1 ValArgSerGlyAsnThrSerArgCys 9
Db 428 GTCCGATCGGGAACACGAGTCGTGT 454

RESULT 2

BJ168506/c

LOCUS BJ168506 688 bp mRNA linear EST 16-OCT-2003
DEFINITION gametophores Physcomitrella patens subsp. patens cDNA clone
DpH18h20 3', mRNA sequence.

BJ168506

BJ168506.1 GI:18336484

EST.

Physcomitrella patens subsp. patens

Physcomitrella patens subsp. patens

Eukaryota; Viridiplantae; Streptophyta; Bryophyta;

Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.

1 (bases 1 to 688)

Nishiyama,T., Fujita,T., Shin-i,T., Seki,M., Nishide,H.,

Uchiyama,I., Kamiya,A., Carninci,P., Hayaishizaki,Y., Shinozaki,K.,

Kohara,Y. and Hasebe,M.

Comparative genomics of Physcomitrella patens gametophytic

transcriptome and Arabidopsis thaliana: implication for land plant

evolution

Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)

JOURNAL MEDLINE

22709184

12808149

CONTACT: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tehini@genes.nig.ac.jp

A backbone of the vector is basically from pBluescript II (KS), that was in vivo excised from a 1-FIC phage vector (Carninci et al. 2001). 5' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA tail was ligated to BamHI site of the

vector(5'- gagAgAgAgggATCCACCTggAgAggTTTTTTTTTTTTT-3' was

used as a lag 3' primer, and

5'-ggTTTCgAgTCATcGTgTTCAGACgCgATgACTCgAGACcGgNNNN-3' as 2nd

5'-hairpin primer, giving the following 5' boarder sequence,

AGCCAAATCGCCAGCTCGAATTCGTCGAGAACCG). cDNA instert could be

amplified with conventional T7 and T3 primers. This full-length

cDNA library was generated according to the method described in

Nishiyama et al. (2003).

Protonemata were blended by the POLYTRON, and then cultivated on

the BCDATG medium for 13- 14 days under the continuous light.

These clones are available from RIKEN Bio Resource Center

(http://www.brc.riken.go.jp/lab/epd/Eng/index.html). The database

of Physcomitrella EST clones is available at the PHYSCObase

(http://moss.nibb.ac.jp).

FEATURES

source

1. 688

/organism="Physcomitrella patens subsp. patens"

/mol_type="mRNA"

/sub_species="patens"

/db_xref="taxon:145481"

/clone="pph18h20"

/tissue type="mixture of chloronemata and young

gametophores with 2 to 5 leaves"

/clone_lib="full length cDNA library, chloronemata and

young gametophores"

ORIGIN

Alignment Scores:
Pred. No.: 10.3 Length: 688
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 69.23% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-174 (1-13) x BJ168506 (1-688)

Qy

1 ValArgSerGlyAsnThrSerArgCys 9

Db

612 GTCCGATCGGGAACACGAGTCGTGT 586

RESULT 3

BJ168506

LOCUS

BJ168506

DEFINITION

HS_2163_A1_G04 MF CIT Approved Human Genomic Sperm Library D Homo

sapiens genomic clone Plate=2163 Col=7 Row=M, genomic survey

sequence.

ACCESSION

AQ125006

VERSION

AQ125006.1

GI:3502172

KEYWORDS

GSS.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 393)

REFERENCE

AUTHORS

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

JOURNAL MEDLINE

99380589

PUBMED

10449784

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Sequence Tagged Connector
Plate: 2163 row: M column: 7
Classes: BAC ends

High quality sequence stop: 393.
Location/Qualifiers

FEATURES

source

1..393
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=2163 Col=7 Row=M"
/sex="male"

/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"

ORIGIN

Alignment Scores:
Pred. No.: 72.6 Length: 393
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 61.54% Indels: 0
DB: 8 Gaps: 0

US-09-851-138C-174 (1-13) x AQL25006 (1-393)

Qy 3 SerGlyAanThrSerArgCysTrp 10
|||||
Db 161 AGTGGCAACACTCCAGGTGCTGG 184

RESULT 4

BF611925

LOCUS

DEFINITION BF611925 731 bp mRNA linear EST 14-DEC-2000
cdna clone IMAGE:3556949 5' similar to FR:018757 O18757 PEROXISOMAL
CA-DEPENDENT SOLUTE CARRIER. 1; mRNA sequence.

ACCESSION BF611925

VERSION BF611925.1 GI:11782060

KEYWORDS EST.

SOURCE Xenopus laevis (African clawed frog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
Xenopodinae; Xenopus; Xenopus.

REFERENCE 1 (bases 1 to 731)

AUTHORS Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L.,

Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y.,

Person, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,

Waterston, R. and Wilson, R.

WashU Xenopus EST project, 1999

Unpublished (1999)

Other ESTs: df13b03.xl

Contact: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

Library constructed by N. Garrett, E. ellefroid, and A.M. Zorn

(Wellcome/CRC Institute). DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov

Seq primer: -40RP from Gibco

High quality sequence stop: 524.

Location/Qualifiers

FEATURES

source

1..731
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:3556949"
/tissue_type="egg, subtracted by stage 13-17 animal cap"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Wellcome CRC PRN3 St13 17 egg animal cap"
/note="Vector: pBSRN3; Site 1: NotI; Site 2: EcoRI; CDNAS
were oligo-dT primed and directionally cloned. Staging
according to Nieukoop and Faber. Library is subtracted
and was constructed by N. Garrett, E. Bellefroid, and A.M.
Zorn, (Wellcome/CRC Institute)."

ORIGIN

Alignment Scores:
Pred. No.: 121 Length: 731
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 61.54% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x BF611925 (1-731)

Qy 3 SerGlyAanThrSerArgCysTrp 10
|||||
Db 691 TCTGGGAACACTCAGGTGCTGG 714

RESULT 5

CE452932/c

LOCUS

DEFINITION tigr-gss-dog-17000319301108 Dog Library Canis familiaris genomic,
genomic survey sequence.

ACCESSION CE452932

VERSION CE452932.1 GI:36745116

KEYWORDS GSS.

SOURCE Canis familiaris (dog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 749)

REFERENCE 1

AUTHORS Kirkness, E.F., Hafna, V., Halpern, A.L., Levy, S., Remington, K.,

Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and

Venter, J.C.

The dog genome: survey sequencing and comparative analysis

Science 301 (5641), 1898-1903 (2003)

22875432

MEDLINE 14512627

COMMENT

Contact: Kirkness EF

The Institute for Genomic Research

Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,

Rockville, MD 20850, USA

Tel: 301-838-0200

Fax: 301-838-0208

Email: ekirknes@tigr.org

Classes: shotgun.

Location/Qualifiers

source

1..749

/organism="Canis familiaris"

/mol_type="genomic DNA"

/strains="Standard Poodle"

/db_xref="taxon:9615"

/clone_lib="Dog Library"

/note="Site 1: BstXI; Libraries were prepared from

peripheral blood"

ORIGIN

Alignment Scores:
Pred. No.: 123 Length: 749
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

```

Query Match: 61.54% Indels: 0
DB: 9 Gaps: 0

US-09-851-138C-174 (1-13) x CE452932 (1-749)

QY 2 ArgSerGlyAsnThrSerArgCys 9
   |||||
Db 71 AGAAGTGGGACACACTTCGCTGT 48

RESULT 6
LOCUS CC336597 815 bp DNA linear GSS 16-MAY-2003
DEFINITION OGOAX72TV_ZM_0.7_1.5_KB_Zea_mays_genomic_clone_ZMMBTA0341L24,
genomic survey sequence.
ACCESSION CC336597
VERSION CC336597.1 GI:30806010
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 815)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Reanick,A., Fraser,C.M., Budinan,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
CONSORTIUM Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
    source
        1..815
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone="ZMMBTA0341L24"
            /clone_lib="ZM 0.7_1.5_KB"
            /note="Vector: pCGRK-; Site 1: HincII; 0.7-1.5 kb
            methylation filtered genomic DNA library"
ORIGIN
Alignment Scores:
Pred. No.: 132 Length: 815
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 61.54% Indels: 0
DB: 8 Gaps: 0

US-09-851-138C-174 (1-13) x CC336597 (1-815)

QY 5 AsnThrSerArgCysTrpIlePro 12
   |||||
Db 130 AATACATCTCGATGTTGGATCCG 107

RESULT 7
LOCUS CC336597 894 bp DNA linear GSS 21-AUG-2003
DEFINITION PUFLA51TB_ZM_0.6_1.0_KB_Zea_mays_genomic_clone_ZMMBTA0674J05,
genomic survey sequence.
ACCESSION CC336597
VERSION CC336597.1 GI:34058399
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 894)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Reanick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
CONSORTIUM Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: PUFLA51TB
Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
    source
        1..904
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone="ZMMBTA0674J05"
            /clone_lib="ZM 0.6_1.0_KB"
            /note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0-kb high
            CoT selected genomic DNA library"
ORIGIN
Alignment Scores:
Pred. No.: 142 Length: 894
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 61.54% Indels: 0
DB: 9 Gaps: 0

US-09-851-138C-174 (1-13) x CG167588 (1-894)

QY 5 AsnThrSerArgCysTrpIlePro 12
   |||||
Db 777 AATACATCTCGATGTTGGATCCG 800

RESULT 8
LOCUS CG167590 904 bp DNA linear GSS 21-AUG-2003
DEFINITION PUFLA51TD_ZM_0.6_1.0_KB_Zea_mays_genomic_clone_ZMMBTA0674J05,
genomic survey sequence.
ACCESSION CG167590
VERSION CG167590.1 GI:34058391
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 904)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Reanick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
CONSORTIUM Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: PUFLA51TB
Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
    source
        1..904
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone="ZMMBTA0674J05"
            /clone_lib="ZM 0.6_1.0_KB"
            /note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0-kb high
            CoT selected genomic DNA library"
ORIGIN

```

```

/organism="zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTa0674J05"
/clone_lib="ZM_0.6_1.0 kb"
/notes="Vector: pCR4-TOPO, Site 1: EcoRI; 0.6-1.0 kb high
Cot selected genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.: 143 Length: 904
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 61.54% Indels: 0
DB: 9 Gaps: 0

US-09-851-138C-174 (1-13) x CG167590 (1-904)

QY 5 AenThrSerArgCysTrpIlePro 12
Db 117 AATACATCTCGATGTTGGATCCG 94

RESULT 9
LOCUS CL490252/c 968 bp DNA linear GSS 01-APR-2004
DEFINITION SAIL 534_H11.v3 SAIL Collection Arabidopsis thaliana genomic clone
ACCESSION SAIL_534_H11.v3, genomic survey sequence.
VERSION CL490252
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 968)
AUTHORS Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
Rullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
JOURNAL A high-throughput Arabidopsis reverse genetics system
MEDLINE Plant Cell 14 (12), 2985-2994 (2002)
PUBMED 22356987
COMMENT Contact: Sessions A
Applied Trait Genetics
Syngenta Biotechnology Inc.
3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
Email: allen.sessions@syngenta.com
ABRC Stock Number C8822704; T-DNA left border flanking sequences of
Syngenta Arabidopsis Insertion Library (SAIL) lines are available
through the Arabidopsis Biological Resource Center (ABRC).
Sequences represent a pool of amplified genomic regions and not
single contiguous sequences.
Class: TDNA tagged.

FEATURES
Location/Qualifiers
source
1..968
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="SAIL 534_H11.v3"
/clone_lib="SAIL Collection"
/notes="T-DNA left border sequences were isolated using a
modified TAIL-PCR strategy"

ORIGIN
Alignment Scores:
Pred. No.: 152 Length: 968
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservativeness: 0

/organism="zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTa0674J05"
/clone_lib="ZM_0.6_1.0 kb"
/notes="Vector: pCR4-TOPO, Site 1: EcoRI; 0.6-1.0 kb high
Cot selected genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.: 143 Length: 904
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 61.54% Indels: 0
DB: 9 Gaps: 0

US-09-851-138C-174 (1-13) x CG167590 (1-904)

QY 5 AenThrSerArgCysTrpIlePro 12
Db 117 AATACATCTCGATGTTGGATCCG 94

RESULT 9
LOCUS CL490252/c 968 bp DNA linear GSS 01-APR-2004
DEFINITION SAIL 534_H11.v3 SAIL Collection Arabidopsis thaliana genomic clone
ACCESSION SAIL_534_H11.v3, genomic survey sequence.
VERSION CL490252
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 968)
AUTHORS Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
Rullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
JOURNAL A high-throughput Arabidopsis reverse genetics system
MEDLINE Plant Cell 14 (12), 2985-2994 (2002)
PUBMED 22356987
COMMENT Contact: Sessions A
Applied Trait Genetics
Syngenta Biotechnology Inc.
3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
Email: allen.sessions@syngenta.com
ABRC Stock Number C8822704; T-DNA left border flanking sequences of
Syngenta Arabidopsis Insertion Library (SAIL) lines are available
through the Arabidopsis Biological Resource Center (ABRC).
Sequences represent a pool of amplified genomic regions and not
single contiguous sequences.
Class: TDNA tagged.

FEATURES
Location/Qualifiers
source
1..968
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="SAIL 534_H11.v3"
/clone_lib="SAIL Collection"
/notes="T-DNA left border sequences were isolated using a
modified TAIL-PCR strategy"

ORIGIN
Alignment Scores:
Pred. No.: 152 Length: 968
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservativeness: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 61.54% Indels: 0
DB: 9 Gaps: 0

US-09-851-138C-174 (1-13) x CL490252 (1-968)

QY 3 SerGlyAenThrSerArgCysTrp 10
Db 914 AGCGGTAAACACGTCGAGTGCTGG 891

RESULT 10
LOCUS BJ052637 164 bp mRNA linear EST 29-SEP-2003
DEFINITION BJ052637 NIBB Mochii normalized Xenopus neurula library Xenopus
laevis cDNA clone XL042b24 3', mRNA sequence.
ACCESSION BJ052637
VERSION BJ052637.1 GI:17498683
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE 1 (bases 1 to 164)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadaou Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tschini@genes.nig.ac.jp
The information of this clone is available through the following
URL.
http://xenopus.nibb.ac.jp.
Location/Qualifiers
source
1..164
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL042b24"
/tissue type="whole embryo"
/dev stage="stage 15"
/clone_lib="NIBB Mochii normalized Xenopus neurula
library"

ORIGIN
Alignment Scores:
Pred. No.: 396 Length: 164
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.85% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-174 (1-13) x BJ052637 (1-164)

QY 2 ArgSerGlyAenThrSerArg 8
Db 112 AGATCTGGTAATACATCTAG 132

RESULT 11
LOCUS BF886593 179 bp mRNA linear EST 18-JAN-2001
DEFINITION RC6-TN0073-271000-012-H12 TN0073 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF886593
VERSION BF886593.1 GI:12277689
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE PUBMED COMMENT

20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC6&t2=RC6-TN0073-271000-012-H12&t3=2000-10-27&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 19

High quality sequence stop: 137.

FEATURES source

1..179
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="TN0073"
/note="Organ: testis normal; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Alignment Scores:
Pred. No.: 425 Length: 179
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.85% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-174 (1-13) x BF886593 (1-179)

Qy 3 SerGlyAsnThrSerArgCys 9

Db 29 TCTGGGAACACACGCGATGT 49

RESULT 12

CE460281/c 201 bp mRNA linear EST 26-MAR-2003
LOCUS
DEFINITION 720152 MARC 6BOV Bos taurus cDNA 3', mRNA sequence.

ACCESSION CB460281

VERSION CB460281.1 GI:29266665

KEYWORDS EST.

SOURCE Bos taurus (cow)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

1 (bases 1 to 201)

REFERENCE Smith,T.P.L., Roberts,A.J., Echtenkamp,S.E., Chitko-McKown,C.G.,

TITLE JOURNAL COMMENT

Wray,J.E. and Keele,J.W.
A second set of bovine ESTs from pooled-tissue normalized libraries
Unpublished (2003)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified with cross match v0.990329.
Plate: FQY8073 row: p column: 2
Seq primer: TAGAAGCAGTCAGTCGAGG.

FEATURES source

1..201
Location/Qualifiers
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 6BOV"
/note="Vector: pCDNA3.1; Site.1: EcoRI; Site.2: NotI;
Library made with RNA pooled from multiple tissues
including liver, lung, hypothalamus, pituitary, and
placenta/endometrium."

ORIGIN

Alignment Scores:
Pred. No.: 467 Length: 201
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.85% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-174 (1-13) x CB460281 (1-201)

Qy 7 SerArgCysTrpIleProVal 13

Db 171 TTCAGATGCTGGATCCAGTT 151

RESULT 13

B93974

LOCUS B93974 227 bp DNA linear GSS 25-JUN-1998
DEFINITION CIT-HSP-2163M7, TF CIT-HSP Homo sapiens genomic clone 2163M7,
genomic survey sequence.

ACCESSION B93974

VERSION B93974.1 GI:2976311

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 227)

Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.

Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)

Unpublished (1998)

Other GSSs: CIT-HSP-2163M7.TR

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html

Seq primer: M13-21;


```
FEATURES
  source
    Class: BAC ends.
    Location/Qualifiers
      1..227
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="GDB:7100513"
        /db_xref="taxon:9606"
        /clone="2163M7"
        /sex="Male"
        /cell_type="Sperm"
        /clone_lib="CIT-HSP"
        /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Alignment Scores:
  Pred. No.: 516      Length: 227
  Score: 7.00      Matches: 7
  Percent Similarity: 100.00%      Conservatives: 0
  Best Local Similarity: 100.00%      Mismatches: 0
  Query Match: 53.85%      Indels: 0
  DB: 8      Gaps: 0

US-09-851-138C-174 (1-13) x B93974 (1-227)

QY      4 GlyAenThrSerArgCysTrp 10
      |||||
Db      88 GCGAACACTTCACAGTGCTGG 108

RESULT 14
CAS525295
LOCUS   KS12053A03 KS12 Capsicum annuum cDNA, mRNA linear EST 15-NOV-2002
DEFINITION
ACCESSION   CAPSICUM
VERSION     CAPSICUM
KEYWORDS    EST.
SOURCE      Capsicum annuum
ORGANISM    Capsicum annuum
REFERENCE   1 (bases 1 to 271)
AUTHORS     Lee,S., Kim,S.-Y., Chung,Y.-H., Shin,H.-J., Goh,S.-H., Pai,H.-S.,
            Hur,C.-G. and Choi,D.
TITLE       Generation of Expressed Sequence Tags from Hot Pepper (Capsicum
            annuum L.) and Sequence Analysis in Relation to Hypersensitive
            Response Against Pathogen
JOURNAL     Unpublished (2001)
COMMENT     Contact: Doil Choi
            Genome Research Center and National Center for Genome Information
            Korea Research Institute of Bioscience and Biotechnology
            P.O. Box 115, Yuseong, Taejeon, 305-600, Republic of Korea
            Tel: 82-42-860-4340
            Fax: 82-42-860-4309
            Email: doil@mail.kribb.re.kr
            Plate: 053 row: A column: 03.
FEATURES
  source
    Location/Qualifiers
      1..271
        /organism="Capsicum annuum"
        /mol_type="mRNA"
        /db_xref="taxon:4072"
        /clone_lib="KS12"

ORIGIN
Alignment Scores:
  Pred. No.: 596      Length: 271
  Score: 7.00      Matches: 7
  Percent Similarity: 100.00%      Conservatives: 0
  Best Local Similarity: 100.00%      Mismatches: 0
  Query Match: 53.85%      Indels: 0
  DB: 6      Gaps: 0

US-09-851-138C-174 (1-13) x CAS525295 (1-271)

QY      6 ThrSerArgCysTrpIlePro 12
      |||||
Db      151 ACCAGCCGTTGCTGGATTCCT 171

RESULT 15
BW202238
LOCUS   BW202238 Nori Satoh unpublished cDNA library, cleaving embryo Ciona
DEFINITION
ACCESSION   BW202238
VERSION     BW202238
KEYWORDS    EST.
SOURCE      Ciona intestinalis
ORGANISM    Ciona intestinalis
REFERENCE   1 (bases 1 to 285)
AUTHORS     Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
TITLE       Expressed genes in Ciona intestinalis (2002c)
JOURNAL     Unpublished (2002)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@acidian.zool.kyoto-u.ac.jp.
FEATURES
  source
    Location/Qualifiers
      1..285
        /organism="Ciona intestinalis"
        /mol_type="mRNA"
        /db_xref="taxon:7719"
        /clone="cic1080k15"
        /tissue_type="whole body"
        /dev_stage="cleaving embryo"
        /clone_lib="Nori Satoh unpublished cDNA library, cleaving
            embryo"

ORIGIN
Alignment Scores:
  Pred. No.: 621      Length: 285
  Score: 7.00      Matches: 7
  Percent Similarity: 100.00%      Conservatives: 0
  Best Local Similarity: 100.00%      Mismatches: 0
  Query Match: 53.85%      Indels: 0
  DB: 5      Gaps: 0

US-09-851-138C-174 (1-13) x BW202238 (1-285)

QY      2 ArgSerGlyAenThrSerArg 8
      |||||
Db      220 AGATCGGGAACACTTCTCGT 240

Search completed: March 3, 2005, 21:58:20
Job time : 823.2 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 14:30:42 ; Search time 71.3333 Seconds
(without alignments)
829.870 Million cell updates/sec

Title: US-09-851-138c-190
Perfect score: 10
Sequence: 1 VKSPCATAS 10

Scoring table: OLGPO*
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8723847

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spol_p/US09851138/runat_28022005_120306_21457/app_query.fasta_1.1123
-DB=N_Geneseq_15Dec04 -QWNT=fastap -SUFFIX=olg.rng -MINMATCH=0.1 -LOOPCL=0
-JOEEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdl
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09851138 @CGN 1 1 1418 @runat_28022005_120306_21457 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq_15Dec04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002as:*
7: Geneseq2002bs:*
8: Geneseq2003as:*
9: Geneseq2003bs:*
10: Geneseq2003cs:*
11: Geneseq2003ds:*
12: Geneseq2004as:*
13: Geneseq2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	100.0	447	2	Aat27962 Hepatitis
2	8	80.0	977	2	Aa040417 Sequence
3	8	80.0	34980	6	Abq81848 Bifidobac
4	7	70.0	98	4	Aac89204 Human bra
5	7	70.0	100	8	AcD77595 E. coli K

C	6	7	70.0	321	6	ABQ90498	Abq90498 M. capsul
	7	7	70.0	328	4	AAS28967	Aas28967 CDNA enco
	8	7	70.0	328	4	AAS30196	Aas30196 DNA enco
	9	7	70.0	328	4	ABA06539	AbA06539 Human cDN
	10	7	70.0	328	4	ABK43946	Abk43946 DNA enco
	11	7	70.0	328	5	AAS29715	Aas29715 Human end
	12	7	70.0	328	5	ADM19205	Adm19205 Novel hum
	13	7	70.0	328	6	ABT07802	Abt07802 Novel hum
	14	7	70.0	328	6	ABV83876	Abv83876 Human pol
	15	7	70.0	328	8	ACD01451	AcD01451 Human pol
	16	7	70.0	328	12	ADI54333	Adi54333 CDNA enco
C	17	7	70.0	414	11	ABD17345	Abd17345 Pseudomon
C	18	7	70.0	435	11	ABD01980	Abd01980 Pseudomon
	19	7	70.0	483	11	ABD10596	Abd10596 Pseudomon
C	20	7	70.0	564	11	ABD08458	Abd08458 Pseudomon
	21	7	70.0	588	10	ABZ41849	Abz41849 N. gonorr
	22	7	70.0	702	13	ADS64196	AdS64196 Bacterial
	23	7	70.0	780	6	ABK74938	Abk74938 Bacillus
C	24	7	70.0	783	10	ADH82756	Adh82756 Enterococ
	25	7	70.0	816	13	ADS63814	AdS63814 Bacterial
C	26	7	70.0	933	8	ACA41109	AcA41109 Prokaryot
C	27	7	70.0	933	10	ABZ41984	Abz41984 N. gonorr
	28	7	70.0	933	13	ADS63446	AdS63446 Bacterial
	29	7	70.0	957	11	ABD08382	Abd08382 Pseudomon
	30	7	70.0	981	9	ADA48273	AdA48273 Rice gene
C	31	7	70.0	1035	6	ABQ90499	Abq90499 M. capsul
C	32	7	70.0	1131	13	ADS48348	AdS48348 Bacterial
C	33	7	70.0	1146	11	ABD01931	Abd01931 Pseudomon
C	34	7	70.0	1155	4	AAH84585	Aah84585 E. coli g
	35	7	70.0	1175	13	ADS61013	AdS61013 Bacterial
	36	7	70.0	1367	4	ABK43658	Abk43658 DNA enco
	37	7	70.0	1367	12	ADI54045	Adi54045 CDNA enco
	38	7	70.0	1374	8	ACA45781	AcA45781 Prokaryot
	39	7	70.0	1377	8	ACA42328	AcA42328 Prokaryot
C	40	7	70.0	1401	5	AAS93933	Aas93933 DNA enco
C	41	7	70.0	1401	5	AAS73477	As73477 DNA enco
	42	7	70.0	1428	11	ABD08406	Abd08406 Pseudomon
C	43	7	70.0	1455	5	AAS29608	Aas29608 Human end
	44	7	70.0	1527	13	ADS56145	AdS56145 Bacterial
	45	7	70.0	1602	11	ADN97354	Adn97354 B. lichen

ALIGNMENTS

RESULT 1
AAT27962
ID AAT27962 standard; DNA; 447 BP.
XX
AC AAT27962;
XX
DT 11-MAR-1997 (first entry)
XX
DE Hepatitis C virus type 10a isolate NN98 bases 478-925.
XX
KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
XX PCR; primer; probe; antibody; infection; ss.
XX
OS Hepatitis C virus.
XX
FN WO9613590-A2.
XX
PD 09-MAY-1996.
XX
PF 23-OCT-1995; 95WO-EP004155.
XX
PR 21-OCT-1994; 94EP-00870166.
XX 28-JUN-1995; 95EP-00870076.
XX
(INNO-) INNOGENETICS NV.
XX
PI Maertens G, Stuyver L;
XX
DR WPI; 1996-251460/25.

DR P-PSDB; AAR96551.
 XX Hepatitis C virus poly:nucleic acid unique to unidentified subtype -
 PT used to develop probes and primers for new sub:types and vaccines to
 PT prevent and treat infection.
 XX
 PS Claim 6; Fig 3; 150pp; English.
 XX
 CC The sequences AAT27937-T27989 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f, 4a-
 CC j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l, 3g,
 CC 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
 CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the genome.
 CC This sequence represents nucleotides 478-925 from the HCV type 10a
 CC isolate NE98. The new HCV types were isolated from patients with chronic
 CC HCV from the Benelux countries, France, Cameroon and Vietnam, because of
 CC their aberrant reactivities. The RNA was extracted, cDNA synthesised and
 CC PCR amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions
 CC were sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences. The
 CC sequences were also used to generate the peptides AAR96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect anti
 CC -HCV antibodies, for HCV typing or to prevent HCV infections
 XX
 SQ Sequence 447 BP; 82 A; 130 C; 114 G; 118 T; 0 U; 3 Other;
 Alignment Scores:
 Pred. No.: 0.158 Length: 447
 Score: 10.00 Matches: 10
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-190 (1-10) x AAT27962 (1-447)
 QY 1 VallySerProCysAlaAlaThrAlaSer 10
 DB 265 GTGAAGTCGCGCTGCGCGCGCGCGCTCT 294

RESULT 2
 AQA0417
 ID AQA0417 standard; DNA; 977 BP.
 XX
 AC AQA0417;
 XX
 XX 24-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 10-AUG-1993 (first entry)
 XX
 DE Sequence encoding R08/26 sarnase.
 XX
 KW Guanyl specific ribonuclease; sarnase; RNase T1; barnase; ss.
 XX
 OS Streptomyces aureofaciens; R08/26.
 XX
 FH Key Location/Qualifiers
 FT sig_peptide 149..346
 FT /*tag= a
 FT mat_peptide 347..640
 FT /*tag= b
 FT misc_feature 347
 FT /*tag= c
 FT /note= "cleavage site FspI (not present in sequence,
 FT introduced by site directed mutagenesis of sequence 344-
 FT 349 from CCCGCC to TGGCA)"
 XX
 PN EP537399-A1.
 XX
 PD 21-APR-1993.
 XX
 PF 16-OCT-1991; 91EP-00402767.

XX 16-OCT-1991; 91EP-00402767.
 PR
 XX (PLBZ) PLANT GENETIC SYSTEMS NV.
 PA
 XX Nazarov V, Botterman J, Stanssens P, Sevcik J;
 PI
 XX WPI; 1993-127352/16.
 DR P-PSDB; AAR34220.
 XX
 CC New ribonuclease DNA from Streptomyces aureofaciens - used for disrupting
 PT metabolism, functioning and/or development of selected cells, esp. plant
 PT cells.
 XX
 PS Claim 2; Page 13-14; 25pp; English.
 XX
 CC The mature protein is a guanyl specific ribonuclease. The sarnase has an
 CC intracellular toxicity between that of RNase T1 and barnase. It can be
 CC selectively expressed in cells of tissue of male or female reproduction
 CC organs to produce male sterile or female sterile plants. (Updated on 25-
 CC MAR-2003 to correct PN field.) (Updated on 24-OCT-2003 to standardise OS
 CC field)
 XX
 SQ Sequence 977 BP; 137 A; 383 C; 324 G; 130 T; 0 U; 3 Other;
 Alignment Scores:
 Pred. No.: 31.7 Length: 977
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 80.00% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-190 (1-10) x AQA0417 (1-977)
 QY 3 SerProCysAlaAlaThrAlaSer 10
 DB 949 TCGCCATGCGCAGCGACGCGATCG 972

RESULT 3
 ABQ81848
 ID ABQ81848 standard; DNA; 349980 BP.
 XX
 AC ABQ81848;
 XX
 DT 19-NOV-2002 (first entry)
 XX
 DE Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1104.
 XX
 KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
 KW antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
 KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
 KW rotavirus; food composition; pharmaceutical composition; gene; ds.
 XX
 OS Bifidobacterium longum.
 OS Synthetic.
 XX
 PN BP1227152-A1.
 XX
 PD 31-JUL-2002.
 XX
 PF 30-JAN-2001; 2001EP-00102050.
 XX
 PF 30-JAN-2001; 2001EP-00102050.
 PR
 XX (NEST) SOC PROD NESTLE SA.
 PA
 XX WPI; 2002-668397/72.
 XX
 PT Novel polynucleotide comprising Bifidobacterium genome sequence useful as
 PT a probe or primer for detecting and/or identifying Bifidobacterium longum
 PT in a biological sample.
 XX

PS Disclosure; SEQ ID NO 1104; 80pp; English.

XX The present invention describes a polynucleotide (I) comprising a

CC sequence of a Bifidobacterium genome selected from the nucleotide

CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at

CC least 90% identity or which hybridises with the sequences given in

CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a

CC fusion protein, comprising a sequence selected from 1097 sequences given

CC in ABP65258 to ABP66354 ligated in frame to a polynucleotide encoding a

CC heterologous polypeptide. (I) has antidiarrheic and antibacterial

CC activities, and can be used as an inhibitor of Salmonella. (I) which is

CC a probe) is useful for the detection and/or identification of

CC Bifidobacterium longum in a biological sample. A carrier containing the

CC lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618) can be

CC used for preventing and/or treating diarrhoea brought about by pathogenic

CC bacteria and/or rotavirus. The carrier is a food composition selected

CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented

CC products, ice-creams, fermented cereal based products, milk based

CC powders, infant formula, pet food or a pharmaceutical composition

CC selected from tablets, liquid bacterial suspensions, dried oral

CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.

CC (I) is useful in DNA arrays or chips to carry out analysis of the

CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent

CC Bifidobacterium related nucleotide sequences given in the Sequence

CC Listing from the present invention but not mentioned further within the

CC specification. N.B. The sequence data for this patent is not represented

CC in the printed specification but is based on sequence information

CC supplied by the European Patent Office

XX

SQ Sequence 349980 BP; 69195 A; 106952 C; 106128 G; 67705 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5.1e+03	Length:	349980
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	80.00%	Indels:	0
DB:	6	Gaps:	0

US-09-851-138c-190 (1-10) x ABQ81848 (1-349980)

QY 3 SerProCysAlaAlaThrAlaSer 10

DB 217973 TCACCGTGGCGGCCACAGCATCT 217996

RESULT 4

AAC89204

ID AAC89204 standard; DNA; 98 BP.

AC AAC89204;

XX

DT 07-MAR-2001 (first entry)

XX

DE Human brain T calcium channel alpha 1G subunit gene exon 4.

XX

KW Ion channel; human; brain T calcium channel; alpha 1G subunit;

KW alpha 1I subunit; epilepsy; drug screening; ds.

XX

OS Homo sapiens.

XX

PN WO200070044-A2.

XX

PD 23-NOV-2000.

XX

PF 08-MAY-2000; 2000NO-US012383.

XX

PR 13-MAY-1999; 99US-0134063P.

PR 04-JUN-1999; 99US-0137547P.

XX

PA (UYJO) UNIV JOHNS HOPKINS.

XX

PI Mittman S, Agnew WS;

XX

DR WPI; 2001-031928/04.

XX P-PSDB; AAB50104.

XX

PT Splice variants of the human brain T calcium channel alpha 1G and alpha

PT 1I subunits and genes encoding the subunits, useful as targets for

PT antiepileptic drugs or for testing compounds or compositions useful in

PT treating epilepsy.

XX

XX Claim 5; Page 16; 89pp; English.

XX

CC The present invention provides the protein and coding sequences for the

CC human brain T calcium channel alpha 1G and alpha 1I subunits. The alpha

CC 1G subunit gene (designated CACNA1G) consists of 38 exons, and

CC alternative processing leads to the production of 64 different proteins.

CC The alpha 1I subunit (designated CACNA1I) consists of 37 exons, and 8

CC proteins are produced due to alternative processing. The sequences

CC provided by the invention are useful for screening drugs for use in the

CC treatment of epilepsy

XX

SQ Sequence 98 BP; 18 A; 30 C; 30 G; 20 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	44	Length:	98
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	70.00%	Indels:	0
DB:	4	Gaps:	0

US-09-851-138c-190 (1-10) x AAC89204 (1-98)

QY 3 SerProCysAlaAlaThrAla 9

DB 55 AGTCGGTGTGCTGCGACCGCT 75

RESULT 5

ACD77595/c

ID ACD77595 standard; DNA; 100 BP.

XX

AC ACD77595;

XX

DT 19-SEP-2003 (first entry)

XX

DE E. coli K12 MG1655 biochip probe SEQ ID 8871.

XX

KW Biochip; gene expression; gut; diagnostic; detection; probe; ss.

XX

OS Escherichia coli.

XX

PN EP1260592-A1.

XX

PD 27-NOV-2002.

XX

PF 17-MAY-2001; 2001EP-00112179.

XX

PR 17-MAY-2001; 2001EP-00112179.

XX

PA (MWGB-) MWG-BIOTECH AG.

XX

PI Donner H, Drescher B, Huber A, Weber J;

XX

XX WPI; 2003-241155/24.

XX

PT Biochip containing probes complementary with open reading frames in

PT Escherichia coli K12, useful for detecting gene expression and expression

PT patterns.

XX

PS Claim 3; Page 1384; 2004pp; German.

XX

CC This invention describes a novel biochip comprising probe spots, each

CC containing many identical probes. The probes are nucleotide sequences of

CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at

CC least one includes a segment of at least 20 bases identical with, or

CC complementary to, a segment of an open reading frame (orf) of Escherichia
 CC coli K12. The biochip is used for specific detection of gene expression
 CC in K12 and for determining the gene expression pattern, e.g. for
 CC diagnostic determination of which E. coli strains are present in the gut,
 CC and to determine the effects of e.g. growth media on gene expression. The
 CC biochip provides as comprehensive as possible detection of the K12
 CC genome, with simultaneous analysis of many different genes with a single
 CC device, and comparison of gene expression between K12 and its mutants or
 CC other E. coli strains in a single experiment. Apart from qualitative and
 CC quantitative information about gene expression, it also allows
 CC measurements of population densities for the various strains. The use of
 CC synthetic oligonucleotides for preparation of probes allows free
 CC variation in probe length and ensures high purity (and thus selectivity,
 CC reactivity and reproducibility); also synthetic probes are generally
 CC shorter than probes prepared by polymerase chain reaction. ACD58731 to
 CC ACD81540 represent oligonucleotide probes used with the biochip described
 CC in the invention
 XX
 SQ Sequence 100 BP; 22 A; 30 C; 27 G; 21 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 44.8 Length: 100
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 70.00% Indels: 0
 DB: 8 Gaps: 0

US-09-851-138C-190 (1-10) x ACD77595 (1-100)

Qy 4 ProCysAlaAlaThrAlaSer 10
 |||||
 Db 29 CCATGTGCGCGCGACTGCATCA 9

RESULT 6
 ACD90498/C
 ID ACD90498 standard; DNA; 321 BP.

XX AC ACD90498;
 XX
 XX 01-OCT-2002 (first entry)
 XX
 XX M. capsulatus gene #483 for DNA array.
 XX
 XX Micro array; gene; ds; differential expression; gene expression.
 XX
 XX Methylococcus capsulatus.
 XX
 XX WO200255655-A2.
 XX
 XX 18-JUL-2002.
 XX
 XX 14-JAN-2002; 2002WO-NO000019.
 XX
 XX 12-JAN-2001; 2001NO-00000235.
 XX
 XX 12-JAN-2001; 2001NO-00000239.
 XX
 XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
 XX (TIGR-) TIGR.

XX Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
 XX Lillhaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
 XX Salzberg SL;
 XX
 XX WPI; 2002-557818/59.

XX Novel DNA array useful for determining differential expression of
 PT Methylococcus capsulatus genes, comprises polynucleotides or
 PT oligonucleotides representative for a selective number of Methylococcus
 PT capsulatus genes.
 XX
 XX Claim 14; Page 271; 678pp; English.

CC The invention relates to a novel DNA array giving a representation of a
 CC number of Methylococcus capsulatus genes. The method of the invention is
 CC useful for determination of the differential expression of the genes of
 CC M. capsulatus, and for studying gene expression on a genomic scale and in
 CC gene expression assays of M. capsulatus genes. The sequences shown in
 CC ACD90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
 CC invention
 XX
 SQ Sequence 321 BP; 54 A; 102 C; 119 G; 46 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 123 Length: 321
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 70.00% Indels: 0
 DB: 6 Gaps: 0

US-09-851-138C-190 (1-10) x ABQ90498 (1-321)

Qy 4 ProCysAlaAlaThrAlaSer 10
 |||||
 Db 188 CCTTGGCGCGCGACCGCTCC 168

RESULT 7
 AAS28967
 ID AAS28967 standard; cDNA; 328 BP.

XX AC AAS28967;
 XX
 XX 21-NOV-2001 (first entry)
 XX
 XX cDNA encoding for human uterine motility-association polypeptide #32.
 XX
 XX Human; uterine motility-association disorder; uterus; pregnancy; labour;
 XX menstrual cycle; gene therapy; ss.
 XX
 XX Homo sapiens.
 XX
 XX WO200155201-A1.
 XX
 XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001317.
 XX
 XX 31-JAN-2000; 2000US-0179065P.
 XX 04-FEB-2000; 2000US-0180628P.
 XX 24-FEB-2000; 2000US-0184664P.
 XX 02-MAR-2000; 2000US-0186350P.
 XX 16-MAR-2000; 2000US-0189874P.
 XX 17-MAR-2000; 2000US-0190076P.
 XX 18-APR-2000; 2000US-0198123P.
 XX 19-MAY-2000; 2000US-0205515P.
 XX 07-JUN-2000; 2000US-0209467P.
 XX 28-JUN-2000; 2000US-0214886P.
 XX 30-JUN-2000; 2000US-0215135P.
 XX 07-JUL-2000; 2000US-0216647P.
 XX 07-JUL-2000; 2000US-0216880P.
 XX 11-JUL-2000; 2000US-0217487P.
 XX 11-JUL-2000; 2000US-0217496P.
 XX 14-JUL-2000; 2000US-0218290P.
 XX 26-JUL-2000; 2000US-0220963P.
 XX 26-JUL-2000; 2000US-0220964P.
 XX 14-AUG-2000; 2000US-0224518P.
 XX 14-AUG-2000; 2000US-0224519P.
 XX 14-AUG-2000; 2000US-0225213P.
 XX 14-AUG-2000; 2000US-0225214P.
 XX 14-AUG-2000; 2000US-0225266P.
 XX 14-AUG-2000; 2000US-0225267P.
 XX 14-AUG-2000; 2000US-0225268P.
 XX 14-AUG-2000; 2000US-0225270P.
 XX 14-AUG-2000; 2000US-0225447P.
 XX 14-AUG-2000; 2000US-0225757P.

```
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232196P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236369P.
PR 02-OCT-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0251799P.
PR 06-DEC-2000; 2000US-0251799P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI, 2001-488777/53.
XX P-PSDB; AAU18125.
XX
XX New nucleic acid molecules encoding 49 human secreted proteins for
XX diagnosing, preventing, treating or ameliorating medical conditions and
XX used as food additives or preservatives.
XX
XX Claim 4; SEQ ID NO 42; 524pp; English.
XX
XX The present invention relates to the isolation of novel human uterine
XX motility-association polypeptides (AAU18094-AAU18152), and cDNA and
XX genomic sequences encoding for these polypeptides. The sequences of the
XX invention are useful in the diagnosis, treatment, prevention and/or
XX prognosis of diseases associated with uterine motility such as pregnancy
XX and labour, and menstrual disorders. The polynucleotide sequences of the
XX invention are also useful in gene therapy. AAS28936-AAS28994 represent
XX cDNA sequences encoding for novel human uterine motility-association
XX polypeptides. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 328 BP; 50 A; 87 C; 99 G; 88 T; 0 U; 4 Other;

Alignment Scores:
Pred. No.: 125 Length: 328
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-190 (1-10) x AAS28967 (1-328)
```

QY 3 SerProCysAlaAlaThrAla 9
 Db 172 AGTCGCTGCTGCGACGCT 192
 RESULT 8
 AAS30196
 ID AAS30196 standard; cDNA; 328 BP.
 XX AC AAS30196;
 XX DT 21-NOV-2001 (first entry)
 XX DE DNA encoding renal and cardiovascular-associated protein, Seq ID 42.
 XX KW Human; antiinflammatory; neuroprotective; immunomodulator; vulnerary;
 KW cardiovascular; cytostatic; nephrotropic; antianaemic; nephritis;
 KW immunosuppressive; kidney disorder; renal failure; hypertension;
 KW cardiovascular disorder; myocardial infarction; blood disorder; anaemia;
 KW blood coagulation disorder; electrolyte imbalance disorder; cancer;
 KW hyponatraemia; hyperkalaemia; neoplastic disorder; nephroma;
 KW autoimmune disease; inflammatory disease; reproductive system disorder;
 KW endocrine disorder; neural activity; neurological disorder;
 KW wound healing; respiratory disorder; ss.
 XX OS Homo sapiens.
 XX FN WO200155328-A2.
 XX PD 02-AUG-2001.
 XX PF 17-JAN-2001; 2001WO-US001359.
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225367P.
 PR 14-AUG-2000; 2000US-0225477P.
 PR 14-AUG-2000; 2000US-0225478P.
 PR 14-AUG-2000; 2000US-0225575P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226688P.
 PR 23-AUG-2000; 2000US-0227182P.
 PR 30-AUG-2000; 2000US-0227009P.
 PR 01-SEP-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 13-OCT-2000; 2000US-0239335P.
 PR 13-OCT-2000; 2000US-0239337P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.

PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-488787/53.
XX P-PSDB; AAU18675.

XX New polynucleotides and polypeptides, useful for diagnosing, treating,
PT preventing or prognosing e.g. kidney, cardiovascular, blood, electrolyte
PT imbalance or neoplastic disorders, autoimmune diseases, cancers.

XX Claim 1; SEQ ID NO 42; 506pp; English.

XX The invention relates to novel nucleic acids and polypeptides useful for
CC diagnosing, treating, preventing and/or prognosing disorders related to
CC these polypeptides. The polynucleotides are especially useful in the
CC diagnosis, prognosis, prevention and/or treatment of diseases which
CC include kidney disorders (e.g. renal failure or nephritis),
CC cardiovascular disorders (e.g. hypertension or myocardial infarction),
CC blood disorders (e.g. anaemia or blood coagulation disorders),
CC electrolyte imbalance disorders (e.g. hyponatraemia or hyperkalaemia),
CC neoplastic disorders (e.g. nephroma or renal cell cancer), autoimmune
CC diseases, cancers, inflammatory diseases, reproductive system disorders,
CC endocrine disorders, neural activity and neurological disorders, wound
CC healing and respiratory disorders. AAS30165-AAS30251 represent the novel
CC human renal and cardiovascular-associated nucleic acid sequences of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at: ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 328 BP; 50 A; 87 C; 99 G; 88 T; 0 U; 4 Other;

Alignment Scores:
Pred. No.: 125 Length: 328
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-190 (1-10) x AAS30196 (1-328)

QY 3 SerProCysAlaAlaThrIla 9
DB 172 AGTCCGTGTGCTGCCACCGT 192

RESULT 9
ABA06539
ID ABA06539 standard; cDNA; 328 BP.
XX

AC ABA06539;
XX 10-JAN-2002 (first entry)
XX Human cDNA SEQ ID NO: 205.
XX Human; gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation; ss.
OS Homo sapiens.
XX WO200154474-A2.
PN 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US001349.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225575P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.

PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.

PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251899P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WFI; 2001-476161/51.
 DR P-PSDB; ABB10317.
 DR
 XX
 PT Isolated nucleic acid molecule encoding an inflammation-associated
 PT polypeptide is used in preventing, treating or ameliorating a medical
 PT condition.
 XX
 XX Claim 1; SEQ ID NO 205; 859pp + Sequence Listing; English.
 XX
 CC The present invention provides human cDNAs, proteins and related genomic
 CC DNAs. These can be used in the treatment of neural, immune system,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal and proliferative disorders and inflammation. The present sequence
 CC is a cDNA of the invention
 XX
 SQ Sequence 328 BP; 50 A; 87 C; 99 G; 88 T; 0 U; 4 Other;
 Alignment Scores:
 Pred. No.: 125 Length: 328
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 70.00% Indels: 0
 DB: 4 Gaps: 0
 US-09-851-138C-190 (1-10) x ABA06539 (1-328)
 QY 3 SerProCysAlaAlaThrAla 9
 ID |||||
 Db 172 AGTCGGTGTGTGCGACCGCT 192
 RESULT 10
 ABK43946
 ID ABK43946 standard; cDNA; 328 BP.
 XX
 AC ABK43946;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 DE DNA encoding novel central nervous system protein #526.
 XX
 KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
 KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
 KW cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;
 KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
 KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
 KW adenocarcinoma; reproductive system disorder; testicular feminisation;
 KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
 KW respiratory disorder; renal disorder; kidney failure; blood disorder;
 KW myocardial infarction; wound healing; cell proliferation; skin aging;
 KW food additive; food preservative; gene therapy; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200155318-A2.
 XX
 PD 02-AUG-2001.
 XX
 XX 17-JAN-2001; 2001WO-US001332.

XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190078P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 30-JUN-2000; 2000US-0214886P.
 PR 28-JUL-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217969P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 23-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 12-SEP-2000; 2000US-0232081P.
 PR 14-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.

PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 05-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 06-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI, 2001-581633/65.

P-FSDB; AAU87616.

New isolated nucleic acid encoding a protein for diagnosing, preventing,

PT treating or ameliorating medical conditions and used as food additives or
PT preservatives.
XX Claim 1; SEQ ID NO 536; 837pp; English.
XX
XX The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical condition and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Alignment Scores:
Pred. No.: 125 Length: 328
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-190 (1-10) x ABK43946 (1-328)

Qy 3 SerProCysAlaAlaThrAla 9

Db 172 AGTCGGTGTCTGGACCGCT 192

RESULT 11

AAS29715

ID AAS29715 standard; cDNA; 328 BP.

XX AC AAS29715;

XX DT 21-NOV-2001 (first entry)

XX DE Human endocrine polypeptide encoding cDNA SEQ ID No 215.

XX Endocrine protein; human; mouse; rabbit; goat; horse; food additive; cat;
KW dog; chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnerrary; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ss;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
KW tissue regeneration; anti-infertility.

XX OS Homo sapiens.

XX PN WO20015364-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001308.

XX

PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-020515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226682P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.

```
PT 02-OCT-2000; 2000US-0237037P.
XX 02-OCT-2000; 2000US-0237038P.
PS 02-OCT-2000; 2000US-0237039P.
XX 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451936/48.
XX P-PSDB; AAU18486.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
disorders of the endocrine system such as reproductive disorders,
```

```
endocrine cancers and also for testing and detection e.g. diagnosis.
Claim 1; SEQ ID NO 215; 604pp; English.
Sequences AAS29511-AAS29736 represent cDNA molecules, which encode the
endocrine polypeptides of the invention. Endocrine polypeptides and their
associated polynucleotides of the invention are useful in the diagnosis,
treatment and prevention of various types of disorders in e.g. humans,
mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
pathological condition can be determined by determining the presence or
absence of a mutation in an endocrine polynucleotide. The treatable
disorders include autoimmune diseases such as rheumatoid arthritis,
hyperproliferative disorders such as neoplasms of the breast or liver,
cardiovascular disorders such as cardiac arrest, cerebrovascular
disorders such as cerebral ischaemia, nervous system disorders such as
Alzheimer's disease, infections caused by bacteria, viruses and fungi,
ocular disorders such as corneal infection, endocrine disorders such as
premature labour and infertility, gastrointestinal disorders such as
Crohn's disease, renal disorders such as glomerulonephritis and
respiratory disorders such as asthma. The polypeptides can also be used
to aid wound healing, to prevent skin aging due to sunburn, to maintain
organs before transplantation, to regenerate tissues and in chemotaxis.
The polypeptides can also be used as a food additive or preservative to
increase or decrease storage capabilities. Note: The sequence data for
this patent did not form part of the printed specification, but was
obtained in electronic format directly from WIPO at
Alignment Scores:
Pred. No.: 125 Length: 328
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 5 Gaps: 0
US-09-851-138C-190 (1-10) x AAS29715 (1-328)
QY 3 SerProCysAlaAlaThrAla 9
DB 172 AGTCGGTGTGCTGGCGCGCT 192
RESULT 12
ADM19205
ID ADM19205 standard; cDNA; 328 BP.
XX
AC ADM19205;
XX
DT 20-MAY-2004 (first entry)
XX
DE Novel human channel/transporter gene #2.
KW ds; gene; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neutropic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; gene therapy; channel/transporter protein;
KW rheumatoid arthritis; neoplasm; cardiac arrest; cerebrovascular disorder;
KW cerebral ischemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;
KW epithelial cell proliferation; skin aging; sunburn; transplantation;
KW chemotaxis; food additive.
XX
OS Homo sapiens.
XX
PN WO200154472-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001307.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
```

PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226686P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 12-SEP-2000; 2000US-0232081P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 21-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0233423P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 23-SEP-2000; 2000US-0234399P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.

PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0244647P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0251989P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 05-JAN-2001; 2001US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-476159/51.

P-PSDB; ADM19684.

Isolated nucleic acid molecule encoding a channel/transporter protein is used in preventing, treating or ameliorating a medical condition.

Claim 1; SEQ ID NO 12; 809pp; English.

The invention relates to an isolated nucleic acid molecule encoding a

CC channel/transporter protein or sequences at least 95% identical to a
CC these. The nucleic acids and proteins encoded by them are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. The antibodies to the proteins can also be used
CC in alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities. This sequence corresponds to a gene of the
CC invention.

SQ Sequence 328 BP; 50 A; 87 C; 99 G; 88 T; 0 U; 4 Other;

Alignment Scores:
Pred. No.: 125 Length: 328
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 5 Gaps: 0

US-09-851-138C-190 (1-10) x ADM19205 (1-328)

Qy 3 SerProCysAlaAlaThrAla 9

Db 172 AGTCGTGCTGCACCGCT 192

RESULT 13

ID ABT07802

AC ABT07802; standard; DNA; 328 BP.

XX ABT07802;

XX 14-NOV-2002 (first entry)

DE Novel human nucleic acid SEQ ID No 42.

XX Immunostimulant; antirheumatic; antiarthritic; neuroprotective;
KW antiallergic; antidiabetic; antiaesthetic; antiinflammatory; nootropic;
KW immunosuppressive; anticoagulant; thrombolytic; antithrombotic;
KW cytostatic; nephrotropic; antiparkinsonian; gynecological; virucide;
KW antibacterial; antiarrhythmic; fungicide; HGFAT05; HMAAE95; HTNBM01;
KW immunodeficiency; autoimmune disorder; allergic reaction; cardiovascular;
KW inflammatory condition; graft-versus-host disease; reproductive system;
KW blood-related disorder; hyperproliferative; endocrine; neurological;
KW respiratory; renal; infectious disease; gastrointestinal; gene therapy;
KW neuronal growth; neuronal disorder; neuro-degenerative condition;
KW keratinocyte growth; human; ds.

XX Homo sapiens.

XX US2002086330-A1.

XX 04-JUL-2002.

XX 17-JAN-2001; 2001US-00764893.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 28-JUN-2000; 2000US-0214886P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0228868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251858P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.

(ROSE/) ROSEN C A.

(RUBE/) RUBEN S M.

(BARA/) BARASH S C.

Rosen CA, Ruben SM, Barash SC;

WPI; 2002-665432/71.

Novel polypeptide useful for diagnosis, prognosis, prevention, and treatment of immune, hyperproliferative, renal, respiratory, cardiovascular, reproductive, endocrine, gastrointestinal and neurological disorders.

Disclosure; Page 229; 335pp; English.

The invention relates to an isolated polypeptide comprising a sequence at least 90% identical to a full length protein sequence selected from 55 sequences given in the specification such as a sequence of 163, 74 or 140 amino acids fully defined in the specification, or the encoding sequence contained in 49 cDNA clones given in specification e.g. HGFAT05, HMAAE95 or HTNBM01. The protein and its encoding nucleic acid are useful for diagnosing a pathological condition or susceptibility to a pathological condition in a subject and for preventing, treating or ameliorating a medical condition. The protein, its encoding nucleic acid and an isolated antibody that can bind to the protein are useful in treating, preventing, diagnosing and/or prognosis immunodeficiencies, autoimmune disorders, allergic reactions and conditions, inflammatory conditions, graft-versus-

CC host disease, blood-related disorders, hyperproliferative disorders,
 CC renal disorders, cardiovascular disorders, respiratory disorders,
 CC neurological disorders, endocrine disorders, reproductive system,
 CC disorders, infectious diseases, and gastrointestinal disorders. The
 CC protein of the invention is useful to stimulate neuronal growth and to
 CC treat, prevent, and/or diagnose neuronal damage which occurs in certain
 CC neuronal disorders or neuro-degenerative conditions, for stimulating
 CC keratinocyte growth, to prevent hair loss, to modulate mammalian
 CC characteristics such as body height, weight, hair color, and to increase
 CC or decrease storage capabilities, fat content, lipid, protein,
 CC carbohydrate, vitamins, minerals, cofactors or other nutritional
 CC components. The nucleic acid of the invention can be used in gene
 CC therapy. This polynucleotide sequence represents one of the novel nucleic
 CC acids of the invention

SQ Sequence 328 BP; 50 A; 87 C; 99 G; 88 T; 0 U; 4 Other;

Alignment Scores:

Pred. No.:	125	Length:	328
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	70.00%	Indels:	0
DB:	6	Gaps:	0

US-09-851-138C-190 (1-10) x ABT07802 (1-328)

Qy 3 SerProCysAlaAlaThrAla 9

Db 172 AGTCCGTGCTGCGACCGCT 192

RESULT 14

ABV83876

ID ABV83876 standard; cDNA; 328 BP.

XX AC ABV83876;

DT 09-DEC-2002 (first entry)

XX DE Human polynucleotide SEQ ID NO 205.

XX Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
 KW antiparkinsonian; antisickling; antianemic; antiarthritic; cancer;
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
 KW gene; ss.

XX OS Homo sapiens.

XX US2002090672-A1.

XX 11-JUL-2002.

XX PF 17-JAN-2001; 2001US-00764853.

XX PR 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 28-JUN-2000; 2000US-0214886P.

PR 07-JUL-2000; 2000US-0216247P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 08-SEP-2000; 2000US-0229513P.
 PR 21-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 25-SEP-2000; 2000US-0234274P.
 PR 27-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239835P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.

XX (ROSE/) ROSEN C A.

PA (RUBE/) RUBEN S M.

PA (BARA/) BARASH S C.

XX Rosen CA, Ruben SM, Barash SC;

XX WPI; 2002-681727/73.

XX P-PSDB; ABP66904.

XX Novel polypeptide useful for diagnosis, prognosis, prevention, and
 PT treatment of immune, hyperproliferative, renal, respiratory,
 PT cardiovascular, reproductive, endocrine, gastrointestinal and
 PT neurological disorders.

XX Claim 1; SEQ ID NO 205; 369pp + Sequence Listing; English.

XX The invention relates to novel genes (ABV83682-ABV84101) and proteins
 CC (ABP66710-ABP67129) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: the sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_sequences

SQ Sequence 328 BP; 50 A; 87 C; 99 G; 88 T; 0 U; 4 Other;

XX Alignment Scores:

Pred. No.: 125 Length: 328
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-190 (1-10) x ABV83876 (1-328)

QY 3 SerProCysAlaAlaThrAla 9

|||||

DB 172 AGTCGGTGCTGCCACCGCT 192

RESULT 15

ACD01451

ID ACD01451 standard; cDNA; 328 BP.

AC ACD01451;

DT 30-JUL-2003 (first entry)

DE Human polynucleotide #32.

KW Human; gene; ss; kidney disorder; cardiovascular disorder; arrhythmia;
glomerulonephritis; urinary tract infection; chronic nephritis; anaemia;
carcinoid heart disease; endocarditis; blood disorder; thrombosis;
KW haemoglobin abnormality; electrolyte imbalance; neoplastic disorder;
KW cancer; respiratory disorder; acute rhinitis; sinusitis; pharyngitis;
KW neurological disorder; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease.

XX OS Homo sapiens.

XX PN US2003013649-A1.

XX PD 16-JAN-2003.

XX PF 21-NOV-2001; 2001US-00989442.

XX PR 31-JAN-2000; 2000US-0179065P.

XX PR 04-FEB-2000; 2000US-0180628P.

XX PR 24-FEB-2000; 2000US-0184664P.

XX PR 02-MAR-2000; 2000US-0186350P.

XX PR 16-MAR-2000; 2000US-0189874P.

XX PR 17-MAR-2000; 2000US-0190076P.

XX PR 18-APR-2000; 2000US-0198123P.

XX PR 19-MAY-2000; 2000US-0205515P.

XX PR 07-JUN-2000; 2000US-0209467P.

XX PR 28-JUN-2000; 2000US-0214886P.

XX PR 30-JUN-2000; 2000US-0215135P.

XX PR 07-JUL-2000; 2000US-0216647P.

XX PR 11-JUL-2000; 2000US-0216880P.

XX PR 11-JUL-2000; 2000US-0217487P.

XX PR 11-JUL-2000; 2000US-0217496P.

XX PR 14-JUL-2000; 2000US-0218290P.

XX PR 26-JUL-2000; 2000US-0220963P.

XX PR 26-JUL-2000; 2000US-0220964P.

XX PR 14-AUG-2000; 2000US-0224518P.

XX PR 14-AUG-2000; 2000US-0224519P.

XX PR 14-AUG-2000; 2000US-0225213P.

XX PR 14-AUG-2000; 2000US-0225214P.

XX PR 14-AUG-2000; 2000US-0225266P.

XX PR 14-AUG-2000; 2000US-0225267P.

XX PR 14-AUG-2000; 2000US-0225268P.

XX PR 14-AUG-2000; 2000US-0225270P.

XX PR 14-AUG-2000; 2000US-0225447P.

XX PR 14-AUG-2000; 2000US-0225757P.

XX PR 14-AUG-2000; 2000US-0225758P.

XX PR 14-AUG-2000; 2000US-0225759P.

XX PR 18-AUG-2000; 2000US-0226279P.

XX PR 22-AUG-2000; 2000US-0226681P.

XX PR 22-AUG-2000; 2000US-0226686P.

XX PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 02-OCT-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.

PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249285P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 06-DEC-2000; 2000US-0256719P.
 PR 08-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 05-JAN-2001; 2001US-0254097P.
 PR 17-JAN-2001; 2001US-0259678P.
 PR 17-JAN-2001; 2001US-00764863.

(ROSE/) ROSEN C A.
 (RUBE/) RUBEN S M.
 (BARA/) BARASH S C.

Rosen CA, Ruben SM, Barash SC;

WPI; 2003-416807/39.
 P-PSDB; ABU97290.

PT New nucleic acids and polypeptides, useful for diagnosing, prognosing,
 PT preventing and/or treating e.g. kidney, cardiovascular, blood,
 PT electrolyte imbalance, neoplastic, respiratory, or neurological diseases
 PT or disorders.

PS Claim 1; Page 227; 363pp; English.

XX The invention relates to isolated nucleic acids encoding novel
 CC polypeptides. The nucleic acids are useful for chromosome mapping, for
 CC radiation hybrid mapping, for detection of cancer, in gene therapy, for
 CC identifying individuals from minute biological samples, as an alternative
 CC to restriction fragment length polymorphism (RFLP) analysis, in forensic
 CC biology and as hybridisation probes for differential identification of
 CC tissues or cell types present in a biological sample. Compositions
 CC comprising the polynucleotides, polypeptides and antibodies specific for
 CC the polypeptides may be used in the diagnosis, prognosis, prevention
 CC and/or treatment of kidney disorders (e.g. glomerulonephritis, urinary
 CC tract infections, chronic nephritis), cardiovascular disorders (e.g.
 CC arrhythmias, carotid heart disease, endocarditis), blood disorders
 CC (e.g. thrombosis, anaemia, haemoglobin abnormalities), electrolyte
 CC imbalance, neoplastic disorders (e.g. cancers), respiratory disorders
 CC (e.g. acute rhinitis, sinusitis, pharyngitis) and neurological disorders
 CC (e.g. Alzheimer's disease, Parkinson's disease, Huntington's disease).
 CC Sequences ACD01420-ACD01491 represent human polynucleotides of the
 CC invention

SQ Sequence 328 BP; 50 A; 87 C; 99 G; 88 T; 0 U; 4 Other;

Alignment Scores:

Pred. No.:	125	Length:	328
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

Query Match: 70.00% Indels: 0
 DB: 8 Gaps: 0

US-09-851-138C-190 (1-10) x ACD01451 (1-328)

Qy 3 SerProCysAlaAlaThrAla 9

Db 172 AGTCGGTGTGCTGCGACCGCT 192

Search completed: March 3, 2005, 16:26:32
 Job time : 93.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:54:32 ; Search time 20.6154 Seconds
(without alignments)
793.716 Million cell updates/sec

Title: US-09-851-138C-190
Perfect score: 10
Sequence: 1 VKSPCAATAS 10

Scoring table:

CGTGG*	
Xgapop 60.0 , Xgapext 60.0	
Fgapop 60.0 , Fgapext 60.0	
Delop 6.0 , Delext 7.0	

Searched: 1202784 seqs, 818138359 residues
Word size: 1

Total number of hits satisfying chosen parameters: 2395490

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09851138/runat_28022005_120307_21495/app_query.fasta_1.1123
-DB=Issued Patents NA -QFWT=fastap -SUFFIX=olig.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -SPART=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09851138 @CGN 1 1 249 @runat_28022005_120307_21495 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	100.0	447	3	US-08-836-075A-51
2	8	80.0	3909	4	US-09-902-540-7652
3	8	80.0	6161	4	US-09-902-540-753
4	7	70.0	414	4	US-09-252-991A-13949
5	7	70.0	417	4	US-09-902-540-3504
6	7	70.0	435	4	US-09-252-991A-584
7	7	70.0	483	4	US-09-252-991A-9200
8	7	70.0	564	4	US-09-252-991A-7062
9	7	70.0	783	4	US-09-134-000C-641
10	7	70.0	837	4	US-09-902-540-4676
11	7	70.0	957	4	US-09-252-991A-6986
12	7	70.0	1146	4	US-09-252-991A-535

c 13	7	70.0	1155	4	US-09-711-164-213	Sequence 213, Appl
14	7	70.0	1428	4	US-09-252-991A-7010	Sequence 7010, Ap
15	7	70.0	1608	4	US-09-902-540-2662	Sequence 2662, Ap
c 16	7	70.0	1629	4	US-09-252-991A-15568	Sequence 15568, A
17	7	70.0	1638	4	US-09-252-991A-494	Sequence 494, App
c 18	7	70.0	1659	4	US-09-252-991A-563	Sequence 563, App
19	7	70.0	1709	3	US-09-594-133-12	Sequence 12, Appl
20	7	70.0	1820	1	US-07-917-111-4	Sequence 4, Appl
21	7	70.0	1820	1	US-08-479-638-4	Sequence 69, Appl
22	7	70.0	1820	1	US-08-294-871A-69	Sequence 69, Appl
23	7	70.0	1820	3	US-08-876-398A-69	Sequence 6397, Ap
c 24	7	70.0	2460	4	US-09-902-540-6397	Sequence 16042, A
c 25	7	70.0	2616	4	US-09-252-991A-16042	Sequence 7118, Ap
c 26	7	70.0	2697	4	US-09-252-991A-16558	Sequence 16558, A
27	7	70.0	2739	4	US-09-902-540-427	Sequence 427, App
28	7	70.0	2856	4	US-09-398-522-51	Sequence 51, Appl
29	7	70.0	3993	4	US-09-426-998-3	Sequence 3, Appl
30	7	70.0	6822	3	US-09-949-016-3859	Sequence 3859, Ap
31	7	70.0	7405	4	US-09-426-998-4	Sequence 4, Appl
32	7	70.0	7741	3	US-09-902-540-1167	Sequence 1167, Ap
c 33	7	70.0	18471	4	US-09-902-540-1123	Sequence 1230, Ap
c 34	7	70.0	19068	4	US-09-902-540-1230	Sequence 1230, Ap
35	7	70.0	24754	4	US-09-949-016-15601	Sequence 15601, A
36	7	70.0	70308	4	US-09-103-840A-2	Sequence 2, Appl
c 37	7	70.0	4403765	3	US-09-103-840A-1	Sequence 1, Appl
c 38	7	70.0	4411529	3	US-09-396-196G-57973	Sequence 57973, A
39	6	60.0	25	4	US-08-530-492-141	Sequence 141, App
40	6	60.0	81	3	US-08-906-517-141	Sequence 15226, A
41	6	60.0	81	3	US-09-621-976-15226	Sequence 2, Appl
c 42	6	60.0	86	4	US-08-487-867-2	Sequence 2, Appl
c 43	6	60.0	115	2	PCT-US96-09358-2	Sequence 2, Appl
c 44	6	60.0	201	1	US-08-470-892-2	Sequence 2, Appl
45	6	60.0	201	1	US-08-470-892-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-836-075A-51
; Sequence 51, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: NABERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; TITLE OF INVENTION: AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870165.9
; FILING DATE: 21 Oct 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:

; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-836-075A-51

Alignment Scores:
Pred. No.: 0.0219 Length: 447
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-851-138C-190 (1-10) x US-08-836-075A-51 (1-447)

Qy 1 VallySerProCysAlaAlaThrAlaSer 10
Db 265 GTGAAGTCGCGCTGCGCGCCACGCTCT 294

RESULT 2

US-09-540-7652/c
; Sequence 7652, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 7652
; LENGTH: 3909
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-7652

Alignment Scores:
Pred. No.: 19 Length: 3909
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-190 (1-10) x US-09-902-540-7652 (1-3909)

Qy 3 SerProCysAlaAlaThrAlaSer 10
Db 2086 AGCCCGTCGCGCTGCCACAGCTTCG 2063

RESULT 3

US-09-902-540-753/c
; Sequence 753, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.

; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 753
; LENGTH: 6161
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-753

Alignment Scores:
Pred. No.: 28.5 Length: 6161
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-190 (1-10) x US-09-902-540-753 (1-6161)

Qy 3 SerProCysAlaAlaThrAlaSer 10
Db 2088 AGCCCGTCGCTGCCACAGCTTCG 2065

RESULT 4

US-09-252-991A-15949/c
; Sequence 15949, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15949
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15949

Alignment Scores:
Pred. No.: 28.3 Length: 414
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-190 (1-10) x US-09-252-991A-15949 (1-414)

Qy 2 LysSerProCysAlaAlaThr 8
Db 69 AAGTCGCGCTGCCAGCTACG 49

RESULT 5

US-09-902-540-3504
; Sequence 3504, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 3504
LENGTH: 417
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-3504

Alignment Scores:
Pred. No.: 28.5 Length: 417
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-190 (1-10) x US-09-902-540-3504 (1-417)

Qy 3 SerProCysAlaAlaThrAla 9
Db 314 TCACCTTGCGCGCAACGGCA 334

RESULT 6

US-09-252-991A-584/c
Sequence 584, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 584
LENGTH: 435
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-584

Alignment Scores:
Pred. No.: 29.6 Length: 435
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-190 (1-10) x US-09-252-991A-584 (1-435)

Qy 4 ProCysAlaAlaThrAlaSer 10
Db 148 CCTTGGCGGCGACGCCAGC 128

RESULT 7

US-09-252-991A-9200
Sequence 9200, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 9200
LENGTH: 483
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9200

Alignment Scores:
Pred. No.: 32.5 Length: 483
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-190 (1-10) x US-09-252-991A-9200 (1-483)

Qy 3 SerProCysAlaAlaThrAla 9
Db 329 TCCCTTGCGCGGACCGCG 349

RESULT 8

US-09-252-991A-7062/c
Sequence 7062, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 7062
LENGTH: 564
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7062

Alignment Scores:
Pred. No.: 37.3 Length: 564
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 4 Gaps: 0

US-09-851-138C-190 (1-10) x US-09-252-991A-7062 (1-564)

Qy 4 ProCysAlaAlaThrAlaSer 10
Db 310 CCTTGGCGGCGACGCCAGC 290

RESULT 9

US-09-134-000C-641/c
Sequence 641, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778


```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 213
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1155)
US-09-711-164-213
```

```
Alignment Scores:
Pred. No.: 71 Length: 1155
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 4 Gaps: 0
```

US-09-851-138C-190 (1-10) x US-09-711-164-213 (1-1155)

```
Qy 4 ProCysAlaAlaThrAlaSer 10
Db 1072 CCATGTGCAGCGACTGCATCA 1052
```

RESULT 14

```
US-09-252-991A-7010
; Sequence 7010, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7010
; LENGTH: 1428
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7010
```

```
Alignment Scores:
Pred. No.: 85.8 Length: 1428
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 4 Gaps: 0
```

US-09-851-138C-190 (1-10) x US-09-252-991A-7010 (1-1428)

```
Qy 4 ProCysAlaAlaThrAlaSer 10
Db 1292 CCCTGGCGGCGACGGCCAGC 1312
```

RESULT 15

```
US-09-902-540-2662
; Sequence 2662, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
```

```
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 2662
; LENGTH: 1608
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-2662
```

```
Alignment Scores:
Pred. No.: 95.4 Length: 1608
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.00% Indels: 0
DB: 4 Gaps: 0
```

US-09-851-138C-190 (1-10) x US-09-902-540-2662 (1-1608)

```
Qy 4 ProCysAlaAlaThrAlaSer 10
Db 668 CCGTGGCGGCAACCGCATCA 688
```

Search completed: March 3, 2005, 22:05:22
Job time : 26.6154 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 3, 2005, 15:43:48 ; Search time 630.154 Seconds
(without alignments)
604.047 Million cell updates/sec

Title: US-09-851-138c-190
Perfect score: 10
Sequence: 1 VKSPCAATAS 10

Scoring table: ~~OMIGOM~~
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Word size: 1

Total number of hits satisfying chosen parameters: 68472384

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: ~~OMIGOM~~
-O=/cgn2_1/USPTO_epool_p/US09851138/runat_28022005_120306_21476/app.query.fasta_1.1123
-DB=EST -QFMT=fastap -SURFIX=olg.rst -MINMATCH=0.1 -LOOPC=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MAIRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09851138 @CGN 1 1 10973 @runat_28022005_120306_21476 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:*

1: gb_est1: *
2: gb_est2: *
3: gb_hcc: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_g881: *
9: gb_g882: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	80.0	193	1 A1595178	----- A1595178 mk09d12.y
2	8	80.0	267	2 BF851443	BF851443 IL5-EN008
3	8	80.0	309	2 BB605338	BB605338 BB605338
4	8	80.0	337	1 AA097387	AA097387 mk09d12.y
5	8	80.0	351	5 BY063673	BY063673 BY063673
6	8	80.0	374	5 BY070345	BY070345 BY070345
7	8	80.0	379	2 AW259762	AW259762 um77h01.y
8	8	80.0	392	6 CA028915	CA028915 HZ63J24r
9	8	80.0	434	1 A1528369	A1528369 ui96f04.y

C	10	8	80.0	476	6	CB335289	CB335289 TC005C05R
	11	8	80.0	479	1	AJ434513	AJ434513
	12	8	80.0	492	6	CA022033	HZ41P11r
C	13	8	80.0	493	7	CK568018	HO08K02w
	14	8	80.0	494	1	AI882167	ul31c09.y
	15	8	80.0	496	1	AI786578	uj17f06.y
	16	8	80.0	498	2	AW260618	um84d09.y
	17	8	80.0	500	6	CD055133	HO08K02s
	18	8	80.0	516	2	BE724505	894076F11
	19	8	80.0	533	1	AI527926	uj30h02.y
	20	8	80.0	551	1	AI048563	ud61d06.y
C	21	8	80.0	558	6	CB336550	TC020F10R
	22	8	80.0	567	5	BU997720	HI08N01r
	23	8	80.0	579	5	BX514185	BX514185
	24	8	80.0	585	5	BX516270	BX516270
	25	8	80.0	590	6	CA020216	HV14K18r
	26	8	80.0	591	6	CA020693	HZ37D05r
	27	8	80.0	605	6	CA020586	HZ36L24r
	28	8	80.0	606	4	BJ479348	BJ479348
	29	8	80.0	611	6	CB868882	HC09L02w
	30	8	80.0	618	6	CB866942	HC05J06w
	31	8	80.0	637	6	CB868956	HC09J13w
	32	8	80.0	660	6	BY739081	BY739081
	33	8	80.0	668	8	BZ406916	OGABH71TM
	34	8	80.0	684	6	BY732790	BY732790
	35	8	80.0	691	6	CB949636	AGENCY
C	36	8	80.0	696	8	BZ708738	OGCDJ17TC
C	37	8	80.0	700	9	CC816170	IO0001P04
	38	8	80.0	701	9	CC820167	IO0001P04
	39	8	80.0	707	6	CB600550	AGENCY
	40	8	80.0	727	2	BF784903	602111060
	41	8	80.0	727	4	BI221101	602939616
	42	8	80.0	737	2	AW160552	au74c12.y
	43	8	80.0	747	8	BZ824408	PUFJR72TB
C	44	8	80.0	747	9	CC568649	QGVFU36TV
	45	8	80.0	752	4	BI144317	602908351

ALIGNMENTS

RESULT 1
A1595178
LOCUS mk09d12.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
DEFINITION IMAGE:492407 5' similar to WP:T04A11.2 CE13124 ;, mRNA sequence.
ACCESSION A1595178
VERSION A1595178.1 GI:4604226
KEYWORDS EST.
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 193)
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Stepoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swallier,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)

TITLE Contact: Marra M/WashU-NCI Mouse EST Project 1999
JOURNAL Washington University School of Medicine
COMMENT 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

This clone is available royalty-free through LINL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the correct orientation)
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand

```

MGI:295855
Seq primer: -40RP from Gibco
High quality sequence stop: 1
POLYA=No.

FEATURES
    source
        1..193
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="IMAGE:492407"
            /dev_stage="19.5 dpc total fetus"
            /lab_host="DH10B (ampicillin resistant)"
            /clone_lib="Soares mouse pJNMF19.5"
            /note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(AT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), sites of a modified pT73 vector
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
52.1	8.00	193	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	80.00%	Indels:	0
DB:	1	Gaps:	0

US-09-851-138c-190 (1-10) x A1595178 (1-193)

```

Qy 1 VallysSerProCysAlaAlaThr 8
    |||||
Db 54 GTGAAGTCCCTTGCGGCACG 77

RESULT 2
BF851443/c
LOCUS
DEFINITION
IL5-EN0086-281100-282-all EN0086 Homo sapiens cDNA, mRNA sequence.
ACCESSION
BF851443.1 GI:12238605
VERSION
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 267)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
66.8	8.00	267	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	80.00%	Indels:	0
DB:	2	Gaps:	0

US-09-851-138c-190 (1-10) x BF851443 (1-267)

```

Qy 3 SerProCysAlaAlaThrAlaSer 10
    |||||
Db 43 TCACCATGTGCGCAACCGCATCA 20

RESULT 3
BF605338/c
LOCUS
DEFINITION
RIKEN full-length enriched, 0 day neonate lung Mus
musculus cDNA clone E030009P10 5', mRNA sequence.
ACCESSION
BF605338
VERSION
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1. (bases 1 to 309)
Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,
Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T.,
Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,
Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y.,
Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,
Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,
Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermotabilization and thermostabilization of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
66.8	8.00	267	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	80.00%	Indels:	0
DB:	2	Gaps:	0

US-09-851-138c-190 (1-10) x BF851443 (1-267)

```

Qy 3 SerProCysAlaAlaThrAlaSer 10
    |||||
Db 43 TCACCATGTGCGCAACCGCATCA 20

RESULT 3
BF605338/c
LOCUS
DEFINITION
RIKEN full-length enriched, 0 day neonate lung Mus
musculus cDNA clone E030009P10 5', mRNA sequence.
ACCESSION
BF605338
VERSION
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1. (bases 1 to 309)
Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T.,
Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T.,
Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,
Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y.,
Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K.,
Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T.,
Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermotabilization and thermostabilization of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
```

Itch,M., Kiteunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5): 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES source
1. 309
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="E030009P10"
/tissue_type="lung"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 0 day neonate lung"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5']
GAGGAGAGCGCGCGCACTCGAGTTTTTTTTTTTNN 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5']
GAGGAGAGATCTCGAGTTAATTAATTAATCCCGCCCCC 3'. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Alignment Scores:
Pred. No.: 74.7 Length: 309
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 2 Gaps: 0

US-09-851-138C-190 (1-10) x BB605338 (1-309)

QY 3 SerProCysAlaAlaThraAlaSer 10

Db 216 TCACCTCGCGCGCTACGCCAGT 193

RESULT 4

AA097387 337 bp mRNA linear EST 25-OCT-1996
LOCUS mk09d12.r1 Soares mouse p3NMf19.5 Mus musculus cDNA clone
DEFINITION IMAGE:492407 5', mRNA sequence.

ACCESSION AA097387

VERSION AA097387.1 GI:1643087

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geiselsberg,K., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,P., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project

TITLE

The WashU-HMI Mouse EST Project

JOURNAL

Unpublished (1996)

COMMENT

WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:295855

Seq primer: -28M13 rev2 from Amer sham

High quality sequence stop: 306.

FEATURES

source

1. 337
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:492407"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares mouse p3NMf19.5"
/note="Vector: p7T3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5']
TGTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTT 3', double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Ronaldo. RNA was kindly provided by Dr. Minoru KO (Wayne State University)."

ORIGIN

Alignment Scores:
Pred. No.: 79.9 Length: 337
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 1 Gaps: 0

US-09-851-138C-190 (1-10) x AA097387 (1-337)

QY 1 VallySerProCysAlaAlaThr 8

Db 88 GTGAAGTCCCTTGTGCGGCCACG 111

RESULT 5

BY063673

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BY063673 351 bp mRNA linear EST 06-DEC-2002
amion Mus musculus cDNA clone 1920029H20 5', mRNA sequence.

BY063673

BY063673.1 GI:26169181

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 351)

Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Baccalo,S., Beisel,K.W., Blake,J.A., Bratt,D., Bruscia,V., Chochoia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,B.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,

Numata,K., Okido,T., Pavan,W.J., Perteau,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Oi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.

TITLE
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL MEDLINE PUBMED
Nature 420, 563-573 (2002)
22354683
12466851

COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Sushiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.jp, URL: <http://genome.gsc.riken.jp/>
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES
source
Location/Qualifiers
1. 351
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I920029H20"
/sex="female"
/tissue_type="amion"
/dev_stage="17 days pregnant adult"
/clone_lib="RIKEN full-length enriched, 17 days pregnant adult female amion"

ORIGIN

Alignment Scores:
Pred. No.: 82.4 Length: 351
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 5 Gaps: 0

US-09-851-138c-190 (1-10) x BY063673 (1-351)

Cy 1 VailysSerProCysalaAlaThr 8

Db 242 GTGAAGTCCCTTGTCGGCCAGC 265

RESULT 6

BY070345

LOCUS

DEFINITION

BY070345

VERSION

BY070345.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Sushiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.jp, URL: <http://genome.gsc.riken.jp/>

Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,

Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,

Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,

Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,

Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and

Hayashizaki,Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gac.riken.go.jp>) for further details.

FEATURES

source Location/Qualifiers

1..374
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="I920072H02"
 /sex="female"
 /tissue_type="amnion"
 /dev_stage="17 days pregnant adult"
 /clone_lib="RIKEN full-length enriched, 17 days pregnant adult female amnion"

ORIGIN

Alignment Scores:
 Pred. No.: 86.5 Length: 374
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservativeness: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 80.00% Indels: 0
 DB: 5 Gaps: 0

US-09-851-138C-190 (1-10) x BY070345 (1-374)

Qy 1 VallysSerProCysAlaAlaThr 8

Db 242 GTGAAGTCCCTTGTGGCCACG 265

RESULT 7

AW259762 379 bp mRNA linear EST 23-DEC-1999
 LOCUS um7h01.y1 Sugano mouse liver mlia Mus musculus cDNA clone
 DEFINITION IMAGE:2317393 5' similar to WP:T0411.2 CE13124 ;, mRNA sequence.
 ACCESSION AW259762
 VERSION AW259762.1 GI:6632743
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 379)
 AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Waller,T., Gibbons,M., Pepe,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
 TITLE The WashU-NCI Mouse EST Project 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:1010037
 High quality sequence stop: 361.
 Seq primer: custom primer used
 Location/Qualifiers

FEATURES

source Location/Qualifiers

1..379
 /organism="Mus musculus"
 /mol_type="mRNA"

/strain="C57BL"
 /db_xref="taxon:10090"
 /clone="IMAGE:2317393"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="Sugano mouse liver mlia"
 /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTITTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAGCTGG and 3' end primer CGACCTCGAGCTGAGCACA."

ORIGIN

Alignment Scores:
 Pred. No.: 87.4 Length: 379
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservativeness: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 80.00% Indels: 0
 DB: 2 Gaps: 0

US-09-851-138C-190 (1-10) x AW259762 (1-379)

Qy 1 VallysSerProCysAlaAlaThr 8

Db 209 GTGAAGTCCCTTGTGGCCACG 232

RESULT 8

CA028915 392 bp mRNA linear EST 24-OCT-2002
 LOCUS HZ63J24r HZ Hordeum vulgare subsp. vulgare cDNA clone HZ63J24
 DEFINITION 5-PRIME, mRNA sequence.
 ACCESSION CA028915
 VERSION CA028915.1 GI:24306879
 KEYWORDS EST.
 SOURCE Hordeum vulgare subsp. vulgare
 ORGANISM Hordeum vulgare subsp. vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.
 1 (bases 1 to 392)
 AUTHORS Radchuk,V., Zhang,H., Weschke,W., Potokina,E. and Wobus,U.
 TITLE Barley ESTs from developing seeds
 JOURNAL Unpublished (2002)
 COMMENT Contact: Stein Nils
 Molecular Markers Group, Department Genbank
 Institute of Plant Genetics and Crop Plant Research (IPK)
 Corrensstr. 3, 06466, Gatersleben, Germany
 Tel: 039482-5522
 Fax: 039482-5595
 Email: stein@ipk-gatersleben.de
 Insert Length: 392 Std Error: 0.00
 Plate: 63 row: 3 column: 24
 Seq primer: M13rev.

Location/Qualifiers
 1..392
 /organism="Hordeum vulgare subsp. vulgare"
 /mol_type="mRNA"
 /cultivar="barke"
 /sub_species="vulgare"
 /db_xref="GABI:282359"
 /db_xref="taxon:112509"
 /clone="HZ63J24"
 /tissue_type="pericarp"

FEATURES

source

/dev_stage="0-7 DAP (days after pollination)"
/lab_host="XL10-Gold"
/clone_lib="HZ"

/dna="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of IMAGE; Site 2: XhoI (3'-end of cDNA); pericarp 0-7 DAP(days after pollination). Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 900 bp"

ORIGIN

Alignment Scores:
Pred. No.: 89.7 Length: 392
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-190 (1-10) x CA028915 (1-392)

QY 3 SerProCysAlaAlaThrAlaSer 10
Db 258 TCACCGTGTGGCCACAGCCTCG 281

RESULT 9
AI528369
LOCUS
DEFINITION
AI528369 434 bp mRNA linear EST 18-MAR-1999
IMAGE:1890271 5' similar to WP:704A11.2 CB13124 ;, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 434)
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

EMAIL: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:974595
Seq primer: custom primer used
High quality sequence stop: 424.
Location/Qualifiers
1. .434
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1890271"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Sugano mouse liver mlia"
/notes="Organ: liver; Vector: pME18S-FL3; Site 1: DraIII (CACTGTGTG); Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer

FEATURES

source
1. .476
/organism="Tribolium castaneum"
/mol_type="mRNA"
/strain="Wild type"
/db_xref="taxon:7070"
/clone="Tc005C05"
/dev_stage="Mixed embryonic stages"
/clone_lib="Tribolium castaneum embryonic cDNA library"
/notes="Vector: pBluescript SK; Site 1: EcoRI; Site 2: XhoI; Uni-ZAP XR cDNA library (Stratagene) constructed by Reinhard Schroder (1995)"

ORIGIN

Alignment Scores:
Pred. No.: 104 Length: 476
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 6 Gaps: 0

[ATGTGGCTTTTITTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCTTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5Kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTGCTCTAAAGCTGG and 3' end primer CGACCTGCAGCTCGAGCACA."

ORIGIN

Alignment Scores:
Pred. No.: 96.9 Length: 434
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 1 Gaps: 0

US-09-851-138C-190 (1-10) x AI528369 (1-434)

QY 1 ValLysSerProCysAlaAlaThr 8
Db 288 GTGAAGTCCCTTGTGCGGCCACG 311

RESULT 10

CB335289/c
LOCUS
DEFINITION
CB335289 476 bp mRNA linear EST 01-JAN-2004
Tc005C05R Tribolium castaneum embryonic cDNA library Tribolium castaneum cDNA clone Tc005C05 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Tribolium castaneum (red flour beetle)
Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenebrionidae; Tribolium.
1 (bases 1 to 476)
Savard,J. and Tautz,D.
A Tribolium castaneum EST project
Unpublished (2003)
JOURNAL
COMMENT

Contact: Savard, J.
Abteilung fur Evolutionsgenetik, AG Tautz
Institut fur Genetik, Universitat zu Koln
Weyertal 121, 50931 Koln, Germany
Tel: 49 221 470 6911
Fax: 49 221 470 5975
Email: savard@uni-koeln.de
Seq primer: M13R.
Location/Qualifiers
1. .476
/organism="Tribolium castaneum"
/mol_type="mRNA"
/strain="Wild type"
/db_xref="taxon:7070"
/clone="Tc005C05"
/dev_stage="Mixed embryonic stages"
/clone_lib="Tribolium castaneum embryonic cDNA library"
/notes="Vector: pBluescript SK; Site 1: EcoRI; Site 2: XhoI; Uni-ZAP XR cDNA library (Stratagene) constructed by Reinhard Schroder (1995)"

FEATURES

source

1. .476
/organism="Tribolium castaneum"
/mol_type="mRNA"
/strain="Wild type"
/db_xref="taxon:7070"
/clone="Tc005C05"
/dev_stage="Mixed embryonic stages"
/clone_lib="Tribolium castaneum embryonic cDNA library"
/notes="Vector: pBluescript SK; Site 1: EcoRI; Site 2: XhoI; Uni-ZAP XR cDNA library (Stratagene) constructed by Reinhard Schroder (1995)"

ORIGIN

Alignment Scores:
Pred. No.: 104 Length: 476
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 6 Gaps: 0

US-09-851-138C-190 (1-10) x CB335289 (1-476)

Qy 2 LysSerProCysAlaAlaThrAla 9
 Db 173 AAATGCCATGTGCAGCAACCGCA 150

RESULT 11

AJ434513 479 bp mRNA linear EST 15-MAR-2002
 LOCUS AJ434513 S00007 Hordeum vulgare cDNA clone S0000700072F08F1, mRNA
 DEFINITION

ACCESSION AJ434513

VERSION AJ434513.1 GI:19522965

KEYWORDS EST.

SOURCE

ORGANISM

Hordeum vulgare
 Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Hordeum.

1 (bases 1 to 479)

REFERENCE Saren,A.-M., Tanskanen,J., Paulin,L. and Schulman,A.H.

AUTHORS Barley EST's

TITLE Unpublished (2002)

JOURNAL Contact: Schulman AH

COMMENT Institute of Biotechnology

University of Helsinki

P.O.Box 56 (Viikinkaari 6A), University of Helsinki FIN-00014,

Finland.

FEATURES Location/Qualifiers

source

1..479

/organism="Hordeum vulgare"

/mol_type="mRNA"

/db_xref="taxon:4513"

/clone="S0000700072F08F1"

/dev_stage="Shoot"

/clone_lib="S00007"

/note="2-,3-,4-days after germination"

ORIGIN

Alignment Scores:
 Pred. No.: 105 Length: 479
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 80.00% Indels: 0
 DB: 1 Gaps: 0

US-09-851-138C-190 (1-10) x AJ434513 (1-479)

Qy 3 SerProCysAlaAlaThrAlaSer 10

Db 74 TCACCGTGTGGCGGCACAGCCTCG 97

RESULT 12

CA022033 492 bp mRNA linear EST 23-OCT-2002
 LOCUS HZ41P11r HZ Hordeum vulgare subsp. vulgare cDNA clone HZ41P11
 DEFINITION 5-PRIME, mRNA sequence.

ACCESSION CA022033

VERSION CA022033.1 GI:24299407

KEYWORDS EST.

SOURCE

ORGANISM

Hordeum vulgare subsp. vulgare
 Hordeum vulgare subsp. vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Hordeum.

1 (bases 1 to 492)

REFERENCE Radchuk,V., Zhang,H., Weschke,W., Potokina,E. and Wobus,U.

AUTHORS Barley ESTs from developing seeds

TITLE Unpublished (2002)

JOURNAL Contact: Stein Nils

COMMENT Molecular Markers Group, Department Genbank

Institute of Plant Genetics and Crop Plant Research (IPK)

Corrensstr. 3, 06466, Gatersleben, Germany

Tel: 039482-5522

Fax: 039482-5595

Email: stein@pk-gatersleben.de

Insert Length: 492 Std Error: 0.00

Plate: 41 row: P column: 11

Seq primer: M13rev.

Location/Qualifiers

1..492

/organism="Hordeum vulgare subsp. vulgare"

/mol_type="mRNA"

/cultivar="barke"

/sub_species="vulgare"

/db_xref="GABI:275438"

/db_xref="taxon:112509"

/clone="HZ41P11"

/tissue_type="pericarp"

/dev_stage="0-7 DAP (days after pollination)"

/lab_host="XL10-Gold"

/clone_lib="HZ"

/note="vector: pBluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); pericarp 0-7 DAP(days after pollination). Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 900 bp"

ORIGIN

Alignment Scores:
 Pred. No.: 107 Length: 492
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 80.00% Indels: 0
 DB: 6 Gaps: 0

US-09-851-138C-190 (1-10) x CA022033 (1-492)

Qy 3 SerProCysAlaAlaThrAlaSer 10

Db 103 TCACCGTGTGGCGGCACAGCCTCG 126

RESULT 13

CK568018/c

LOCUS CK568018

DEFINITION

sequence.

ACCESSION CK568018

VERSION CK568018.1

KEYWORDS EST.

SOURCE

ORGANISM

Hordeum vulgare

Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Hordeum.

1 (bases 1 to 493)

Zierold,U. and Schweizer,P.

Barley ESTs from pathogen-attacked leaf epidermis

Unpublished (2003)

CONTACT: Patrick Schweizer

Transcriptome Analysis, Cytogenetics Department

Institute of Plant Genetics and Crop Plant Research (IPK)

Corrensstr. 3, D-06466 Gatersleben, Germany

Tel: 0049 (0)39482-5660

Fax: 0049 (0)39482-5595

Email: schweiz@pk-gatersleben.de

Insert Length: 493 Std Error: 0.00

Plate: 8 row: K column: 2

Seq primer: 17.

Location/Qualifiers

```

source
1. .493
/organism="Hordeum vulgare"
/mol_type="mRNA"
/cultivar="Ingrid BC mlo-5"
/db_xref="CABI:901955"
/db_xref="taxon:4513"
/clone="H008K02"
/tissue_type="leaf epidermis, 6 h and 24 h post
inoculation with Blumeria graminis"
/dev_stage="7 d after germination"
/lab_host="X110-Gold"
/clone_lib="HO"
/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of
cDNA); Site 2: XhoI (3'-end of cDNA); Approximately 5 % of
the clones correspond to cDNA from the fungi B. graminis
hordei and tritici, respectively. Due to a cloning
artefact caused by the kit, in most cases the EcoRI site
is NOT present, as well as the EcoRI adapter used for
cloning. To excise the insert, restriction sites upstream
EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
due to the cloning system used Blue/white selection for
recombinants is not 100% reliable. Average insert size is
1.2 kb"

ORIGIN
Alignment Scores:
Pred. No.: 107 Length: 493
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 7 Gaps: 0

US-09-851-138C-190 (1-10) x CK568018 (1-493)

QY 3 SerProCysAlaAlaThrAlaSer 10
Db 485 TCTCCCTGCGCGGTACCGCAGC 462

RESULT 14
AI882167 494 bp mRNA linear EST 22-JUL-1999
LOCUS ul31c09.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
DEFINITION IMAGE:2099920 5' similar to WP:T04A11.2 CE13124 ;, mRNA sequence.
ACCESSION AI882167
VERSION AI882167.1 GI:5567256
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 494)
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:996852
Seq primer: custom primer used.
Location/Qualifiers
1. .494
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"

FEATURES
source
1. .493
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"

/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2099920"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Sugano mouse kidney mkia"
/note="Organ: kidney; Vector: pME18S-FL3; Site 1: DraIII
(CACTGTGG); Site 2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCTTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTTGTCTCTAAAGCTGCG and 3' end
primer CGACCTGCAGCTCGACACA."

ORIGIN
Alignment Scores:
Pred. No.: 107 Length: 494
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 80.00% Indels: 0
DB: 1 Gaps: 0

US-09-851-138C-190 (1-10) x AI882167 (1-494)

QY 1 VallySerProCysAlaAlaThr 8
Db 237 GTGAAGTCCCTTGTGCGCCACG 260

RESULT 15
AI786578 496 bp mRNA linear EST 02-JUL-1999
LOCUS uJ17f06.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
DEFINITION IMAGE:1908323 5' similar to WP:R02D1.1 CE12484 ;, mRNA sequence.
ACCESSION AI786578
VERSION AI786578.1 GI:5334294
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 496)
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Other ESTs: uJ17f06.x1
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:976519
Seq primer: custom primer used
High quality sequence stop: 448.
Location/Qualifiers
1. .496
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"

FEATURES
source
1. .496
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"

```



```
/db_xref="taxon:10090"  
/clone="IMAGE:1908323"  
/sex="female"  
/dev_stage="adult"  
/lab_host="DH10B"  
/clone_lib="Sugano mouse kidney mkia"  
/note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII  
(CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA  
was primed with an oligo(dT) primer  
[ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was  
ligated to a DraIII adaptor [TGTGGCCTACTGG], digested  
and cloned into distinct DraIII sites of the pME18S-FL3  
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should  
be used to isolate the cDNA insert. Size selection was  
performed to exclude fragments <1.5kb. Library  
constructed by Dr. Sumio Sugano (University of Tokyo  
Institute of Medical Science). Custom primers for  
sequencing: 5' end primer CTCTGTCTCTAAAGCTGGG and 3' end  
primer CGACCTGCAGCTCGACACA."
```

ORIGIN

Alignment Scores:			
Pred. No.:	107	Length:	496
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	80.00%	Indels:	0
DB:	1	Gaps:	0

US-09-851-138C-190 (1-10) x AT786578 (1-496)

Qy	1	VallysSerProCysAlaAlaThr	8
Db	141	GTGAAGTCCCTTGTGGCCACG	164

Search completed: March 3, 2005, 21:50:25
Job time : 635.154 secs

THIS PAGE BLANK (USPTO)